SREAgent 产品手册

本手册以“前端界面可见功能”为中心组织，面向一线 SRE / 值班同学 / 平台管理员，覆盖从场景化排障到系统治理与扩展能力（Skills）的完整使用路径。

2026-04-15

目录

1 概览

1.1 概述

本手册以“前端界面可见功能”为中心组织，面向一线 SRE / 值班同学 / 平台管理员，覆盖从场景化排障到系统治理与扩展能力（Skills）的完整使用路径。

1.1.1 快速开始

- **我想从“值班排障”开始**：先看 问题看板，再根据来源进入 告警治理 或 集群全景 做深挖。
- **我想把排障流程标准化**：从 变更管理 沉淀 Runbook，再用 定时任务 做周期巡检。
- **我想从“对话诊断”开始**：直接进入 新对话，需要复盘时使用 历史对话 管理资产。

1.1.2 功能导航

1.1.2.1 智能运维

- 问题看板
- 集群全景
- 告警治理
- 变更管理

1.1.2.2 Agent 控制

- 新对话
- 定时任务
- 历史对话

1.1.2.3 空间配置

- AGENTS.md
- Skills 管理
- 成员管理

- 集群配置
- 告警源配置

1.1.2.4 系统设置

- 工作空间管理
- 用户管理
- 通用设置
- 个人中心
- 消耗分析
- 许可证
- 关于

1.1.3 使用前置说明

- 系统通常要求先选定工作空间再使用除登录外的大部分功能。
- 根据角色权限不同，部分功能模块可能不可见或仅支持只读操作。

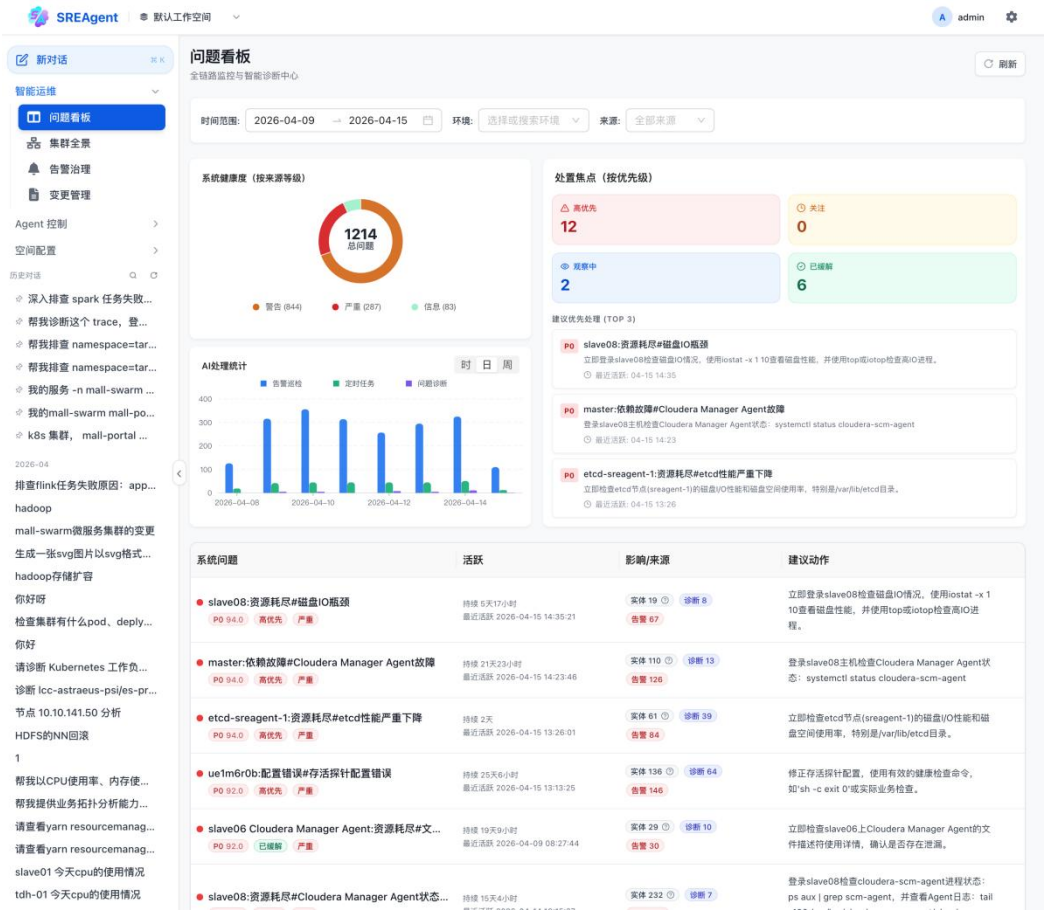
2 智能运维

2.1 问题看板

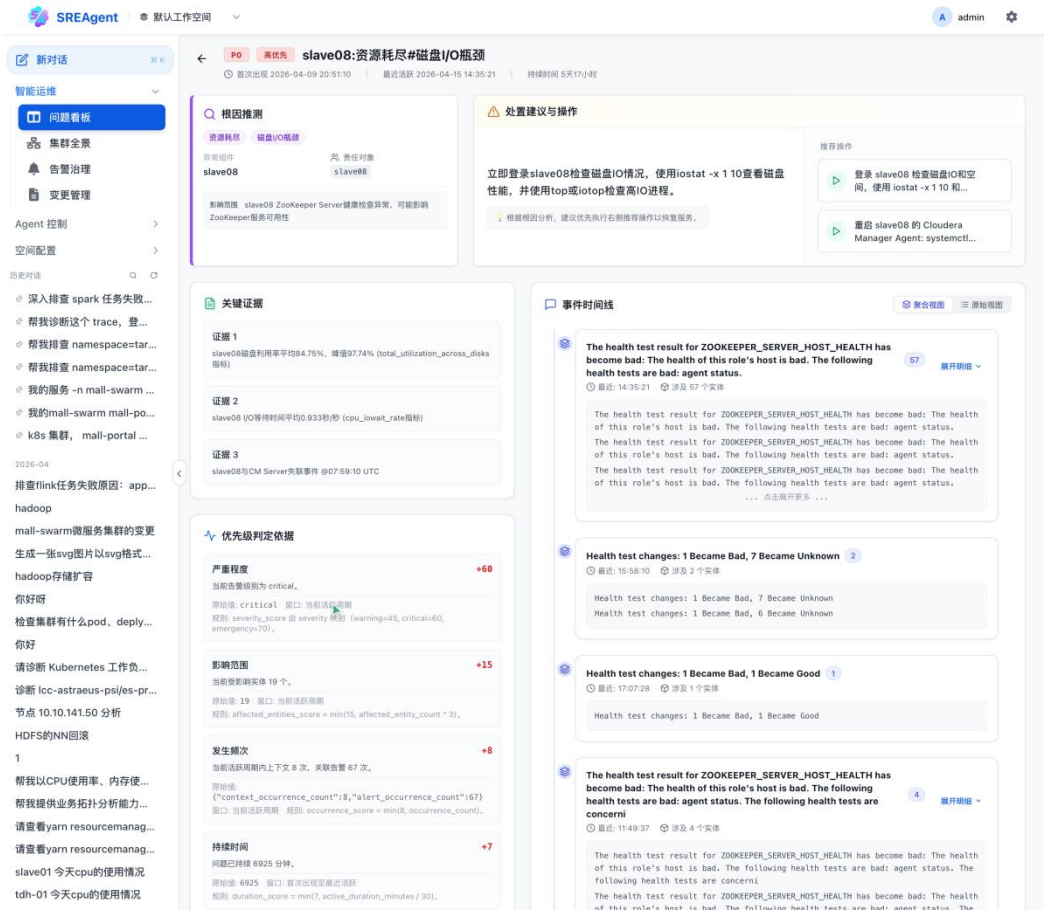
问题看板用于把工作空间内的告警、巡检、对话诊断结果聚合成“可跟踪的问题资产”，帮助值班同学快速判断：哪些问题最紧急、影响面多大、已经走到哪一步，以及下一步该怎么处理。

2.1.1 页面概览（截图）

问题看板总览页：



问题详情页：



2.1.2 入口

- **菜单入口**：侧边栏「问题看板」。

2.1.3 页面构成

问题看板通常由三部分组成：

- **筛选区**：用时间、环境/集群、来源快速定位关注范围。
- **概览区**：从健康度、AI 处理量、处置焦点等维度，给出“值班视角”的总览。
- **问题列表**：以“问题”为粒度聚合展示，并提供进入详情、查看来源、分页等能力。

2.1.4 核心操作指南

2.1.4.1 1) 先用筛选收敛范围

常见筛选维度（以界面为准）：

- **时间范围**：支持快捷时间（如最近 7/14/30 天、本周、上周）与自定义区间。
- **环境/集群**：按分组展示，支持搜索，用于把问题限定在某个环境或集群范围内。
- **来源**：按来源类型过滤（告警 / 巡检 / 对话），适合在洪峰时先把“入口类型”拆开处理。

使用建议：

- 交接班优先用“时间范围”拉出最近活跃问题，再按“来源”拆分责任路径。
- 大盘噪声多时先选定关键环境/集群，避免被非关键环境干扰判断。

2.1.4.2 2) 读懂概览区：健康度与处置焦点

概览区常见卡片（以界面为准）：

- **系统健康度（按来源等级）**：按告警等级/严重性聚合展示占比与数量；无异常时会提示“暂无异常问题”。
- **AI 处理统计**：按来源类型聚合展示 AI 处理量，并支持按粒度切换（时/日/周），用于判断“平台处理压力”和“诊断覆盖度”。
- **处置焦点（按优先级）**：把问题按处置优先级聚类，并给出 Top 处理建议入口，点击可直接进入问题详情。

2.1.4.3 3) 使用问题列表定位“该先处理谁”

问题列表中常见信息（以界面为准）：

- **系统问题**：问题标题 + 标签组合（优先级分数、处置状态、严重程度）。标题可点击进入问题详情。
- **活跃度**：展示持续时间（首次出现 ~ 最近活跃）与最近活跃时间，便于识别“长期未解”与“刚爆发”的问题。
- **影响/来源**：
  - 影响：受影响实体数量、诊断数量等摘要。
  - 来源：按类型统计（告警/问题诊断/定时任务等）。当来源支持明细时，可进一步查看来源详情。
- **建议动作**：展示系统给出的下一步建议；没有建议时会提示“暂无建议”。

2.1.4.4 4) 查看来源详情：把“问题”追溯回“触发入口”

当你需要回答“这个问题是怎么被发现的”时，可在列表中打开“来源详情”，常见用途：

- 把问题关联回具体告警内容（用于工单/群同步）。
- 判断是否由巡检批量触发，避免对同一类问题重复投入。
- 从来源入口反向进入对应模块做进一步处理（例如进入告警治理查看告警上下文）。

2.1.4.5 5) 问题详情：从结论到证据到时间线

问题详情页通常遵循“结论先行 + 证据支撑 + 时间线复盘”的阅读顺序：

1. **页头信息**：显示问题名称、优先级/状态标签，以及首次出现/最近活跃/持续时间，帮助你快速判断紧急程度与是否持续。
2. **结论总览**：
    - 根因推测（可能包含根因分类、异常组件、涉及对象等摘要）。
    - 处置建议与推荐操作（可点开查看更完整的 Markdown 操作说明）。
3. **关键证据 & 优先级依据**：
    - 关键证据：支撑结论的证据条目（无数据时会提示暂无证据）。
    - 优先级判定依据：解释为什么被标记为高优先/关注/观察中等。
4. **事件时间线（两种视图）**：
    - **聚合视图**：把相似事件聚合成事件组，便于快速浏览“主要波动点”，并支持展开明细。
    - **原始视图**：逐条按时间展示原始事件，适合复盘关键时刻的细节。

2.1.5 最佳实践

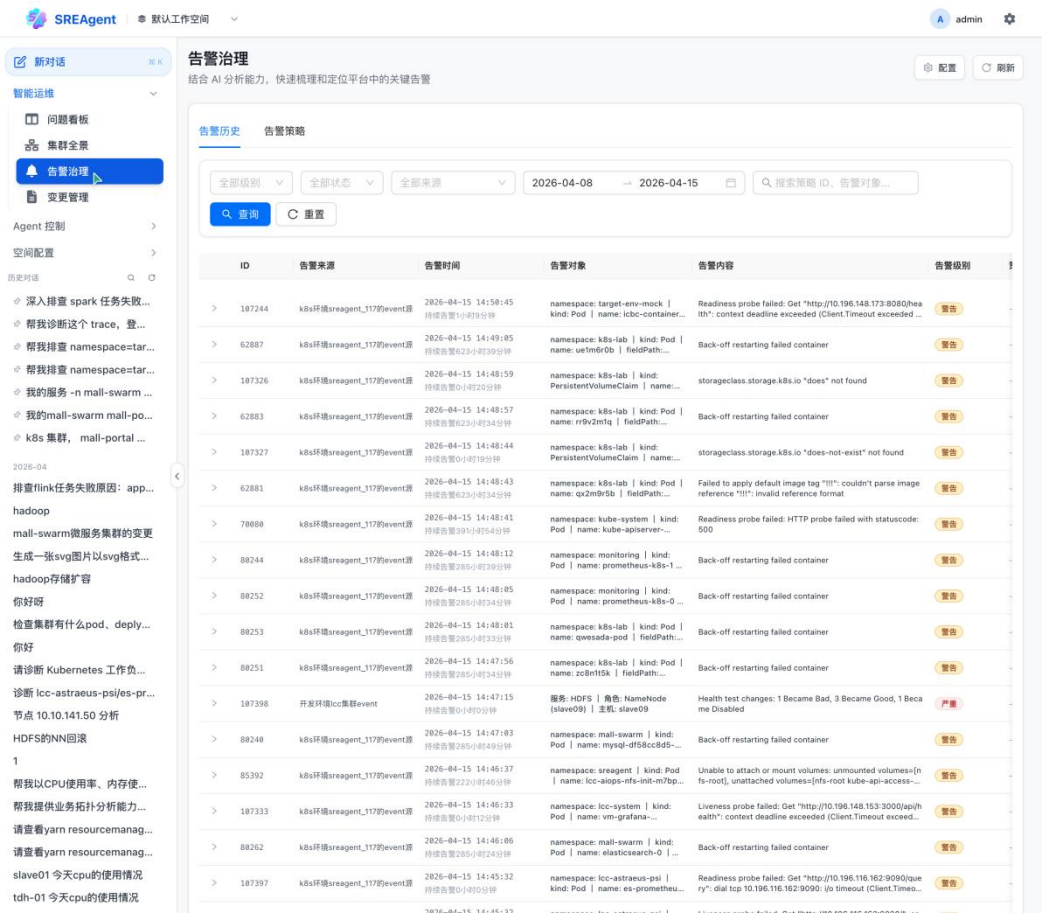
- **先看处置焦点再看列表**：先用 Top 建议定位“最值得投入”的问题，再进入列表做细化筛选。
- **把问题当作资产管理**：一旦进入详情确认结论可靠，就把“根因 + 证据链 + 建议动作”复用到工单与复盘材料里。
- **联动闭环**：问题来自告警就回到告警治理做策略优化；来自巡检就回到定时任务调整巡检口径与频率。

2.2 告警治理

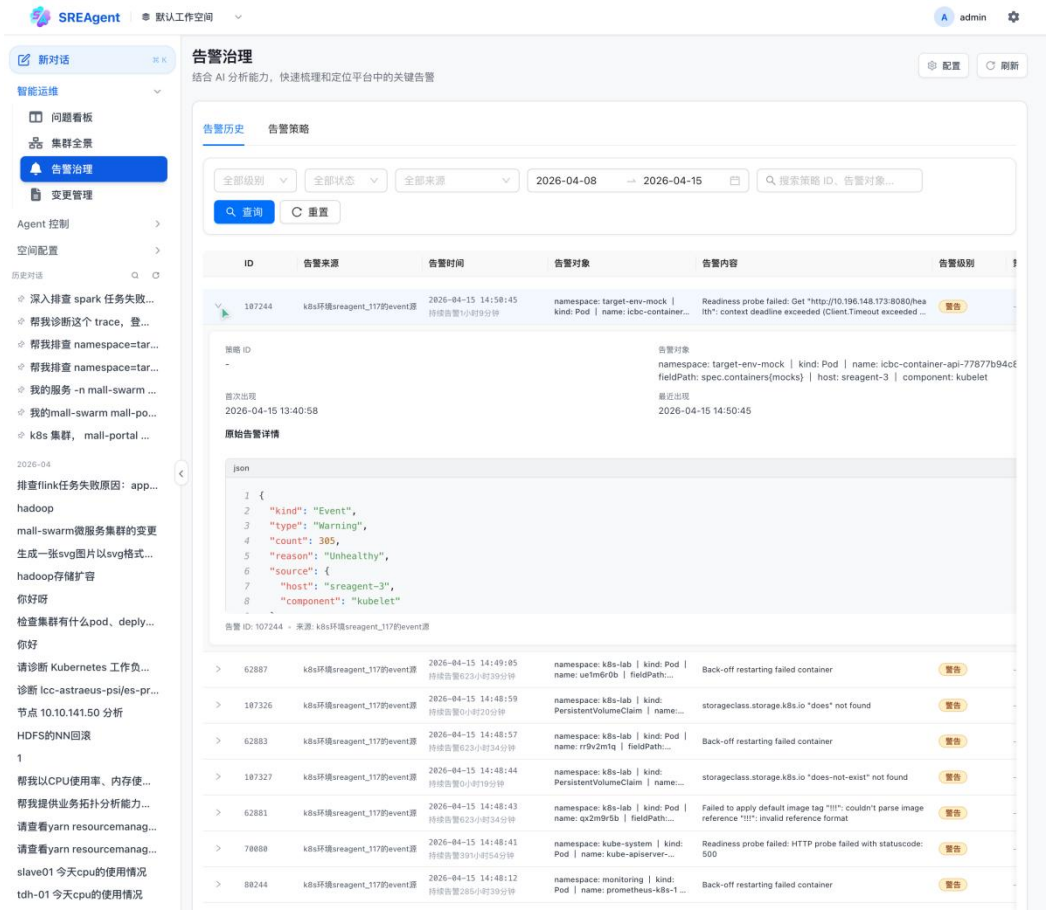
告警治理用于统一展示工作空间内的告警事件，并结合 AI 分析能力，将“告警文本”快速转化为“可执行结论（证据链 + 根因推测 + 建议动作）”。适合值班同学做告警收敛、优先级判断与一键复盘。

2.2.1 页面概览（截图）

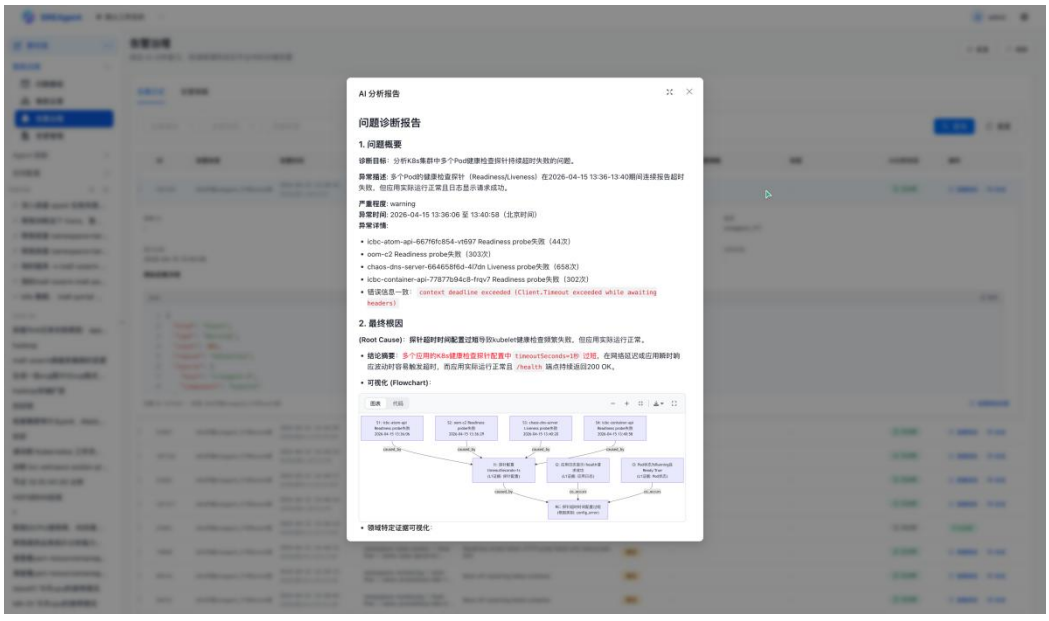
告警历史列表（筛选区 + 列表 + 操作）：



展开告警查看原始上下文（策略 ID/对象/集群/标签/原始告警）：



诊断报告弹窗（Markdown 报告 + 全屏）：



2.2.2 入口

- **菜单入口**：侧边栏「告警治理」。

2.2.3 页面构成

告警治理主要由两部分组成：

- **查询区（筛选/检索）**：用于快速缩小范围，定位“真正需要处理”的告警。
- **告警列表**：用于浏览、展开查看详情，并执行 AI 分析/查看报告/进入会话等动作。

部分版本会提供第二个 Tab（如“告警策略”），用于从“策略维度”管理与回溯告警来源。

2.2.4 核心操作指南

2.2.4.1 1) 筛选与定位告警

常见筛选维度（以界面为准）：

- **告警级别**：信息 / 警告 / 严重 / 紧急。
- **AI 分析状态**：未分析 / 分析中 / 已分析。
- **告警来源**：按告警源（数据源/平台）过滤。
- **时间范围**：按开始/结束时间过滤。
- **关键词**：支持按策略 ID、告警对象等信息检索。

使用建议：

- 交接班/日常巡检优先用“时间范围 + AI 分析状态（未分析）”快速拉出待处理集合。
- 告警洪峰时先按“严重/紧急”筛选，并结合“告警对象”聚类处理，减少切换成本。

2.2.4.2 2) 展开查看告警上下文（非跳转）

在告警列表中点击展开按钮，可在当前列表内查看更多上下文信息，例如：

- 策略 ID 与策略名称（便于定位是否为重复告警、阈值是否合理）。
- 告警对象与集群信息（便于快速判断影响面）。
- 标签（Labels）与原始告警详情（用于补充证据、贴到工单）。

2.2.4.3 3) 触发 AI 分析

当告警处于“未分析”状态时，可点击“AI 分析”：

- 系统会将告警上下文输入给诊断引擎。
- 列表中会展示“分析中”状态，分析完成后可查看报告与进入会话。

建议：

- 对同一策略短时间内的大量重复告警，建议先分析一条代表性告警，再批量处置或调整策略。

2.2.4.4 4) 查看诊断报告

当告警已分析完成后，可点击“查看报告”打开报告弹窗。报告通常包含：

- 现象确认（是否持续、是否波动）。
- 证据链（关键日志/指标/事件）。
- 根因推测（最可能原因与判断依据）。
- 建议动作（立即止血 + 长期治理）。

弹窗常见能力：

- **全屏查看**：适合长报告阅读与复盘。
- **复制内容**：用于粘贴到工单、值班群或复盘文档。

2.2.4.5 5) 进入会话（复盘/继续追问）

当告警支持关联会话时，可以点击“会话”进入对话详情：

- 在对话中可继续追加日志片段、时间范围、集群范围等关键信息。
- 适合把告警分析从“报告”升级为“交互式排障”。

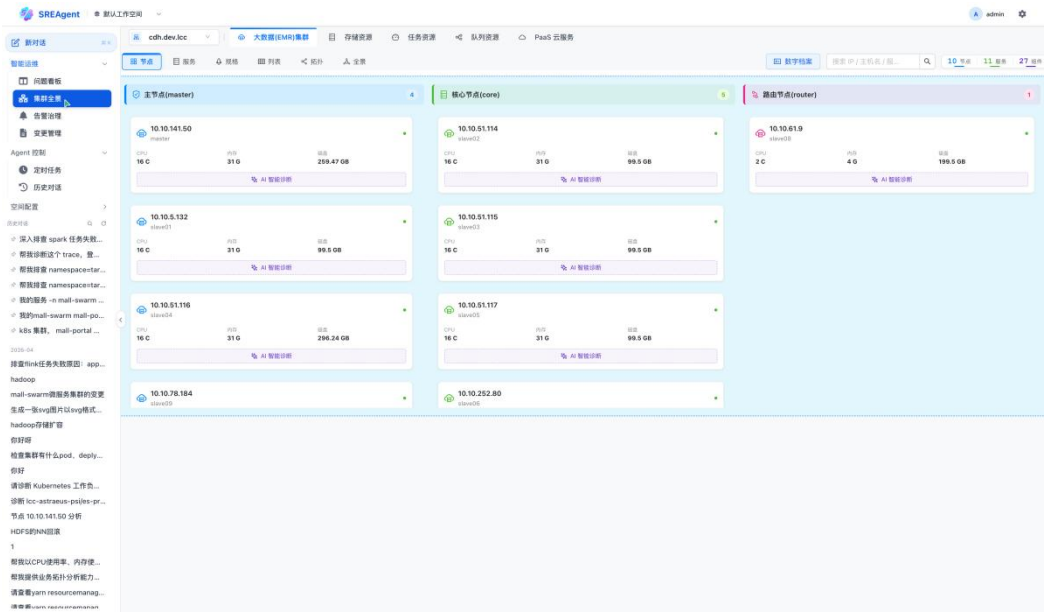
2.2.5 最佳实践

- **先收敛再深挖**：先用筛选把“真正影响业务”的告警收敛出来，再对代表性告警进行 AI 分析与深挖。
- **报告即工单材料**：统一用“根因 + 证据链 + 建议动作”的结构同步，减少往返沟通。
- **策略回溯**：对高频噪声告警，把“告警对象/标签/触发条件”沉淀为策略优化建议，形成治理闭环。

2.3 集群全景

集群全景用于从“集群资源拓扑 + 运行态列表/指标”的视角，帮助值班同学快速理解集群结构、定位资源瓶颈与异常热点，并在需要时一键联动到对话诊断。

2.3.1 页面概览（截图）



2.3.2 入口

- **菜单入口**：侧边栏「集群全景」。

2.3.3 页面构成

集群全景通常具备以下结构（以实际界面为准）：

- **集群选择器**：用于切换当前观测的集群（按分组展示，支持搜索）。
- **子 Tab 区**：根据集群类型与接入能力，展示不同的子域（如大数据/EMR、K8s、存储、任务、队列、PaaS 等）。
- **详情抽屉/弹窗**：用于查看节点、实例、任务等对象的详细信息，并提供进一步操作入口。

提示：页面会把你当前的“集群 + 子 Tab + 子模块”状态写入地址栏，方便把某个视角分享给同事复现问题。

2.3.4 核心概念

- **集群**：来源于工作空间的集群配置；若下拉中没有可选项，请先完成「空间配置 -> 集群配置」维护。
- **子 Tab（一级视角）**：按域划分的观测入口，例如“EMR/K8s/存储/任务/队列/PaaS”。
- **子模块（二级视角）**：在某些子 Tab 内进一步拆分，例如 K8s 的“拓扑/负载/网络”，任务的“列表/概览”等。

2.3.5 核心操作指南

2.3.5.1 1) 选对集群，再选对视角

建议使用顺序：

- 日常巡检：先选定集群，再看“拓扑/概览”类视图确认是否有明显风险。
- 出现慢任务/资源拥堵：直接进入“任务/队列”类视图做定位。
- 存储成本/稳定性：进入“存储”类视图，关注治理项（如小文件/生命周期）。

2.3.5.2 2) 大数据（EMR）视角：从节点到组件到拓扑

常见能力（以界面为准）：

- **视图切换**：支持按节点、服务、规格、表格、架构拓扑等方式查看同一集群。
- **搜索**：支持按 IP/主机名/服务关键字过滤，快速定位目标节点或组件。
- **节点详情抽屉**：点击节点卡片/列表行/拓扑节点后打开，常见内容包括：
  - 基础信息：IP（可复制）、节点组、机型、申请时间、健康/状态标签等。
  - 资源规格：CPU/内存/磁盘汇总与磁盘挂载明细。
  - 组件服务：节点上运行的服务/组件列表与状态。
- **数字档案（如有）**：用于查看集群“数字档案”（基础信息/组件信息等），适合做上线前核对与容量评估。

2.3.5.3 3) K8s 视角：概览 + 拓扑/链路 + 资源清单

K8s 子域通常提供多个子模块（以界面为准）：

- **概览**：展示集群级核心指标汇总。
- **拓扑 / 链路**：以应用/服务为主线展示关系，常配合命名空间与服务关键字过滤使用。
- **负载 / 网络**：提供资源清单式的检索入口，适合定位异常对象。
- **存储（配置）**：面向配置与存储资源的检索入口，支持按命名空间辅助过滤。
- **策略/诊断（如有）**：策略与诊断能力入口，常用于治理型能力接入。

使用建议：

- 先用命名空间缩小范围，再用关键字搜索服务名，减少拓扑视图的噪声。

2.3.5.4 4) 存储视角：容量概览到治理清单

存储子域常见子模块（以界面为准）：

- **存储概览**：容量、文件数、小文件等核心指标概览。
- **数表透视**：以“库表”为主线的透视分析入口，适合做热点表/异常表定位。
- **生命周期管理**：用于查看生命周期策略与相关入口。
- **小文件治理**：提供待治理表清单与治理动作入口，常见能力：
  - 去扫描：对单表触发扫描/更新指标。
  - 建任务：创建合并任务（通常需要补充治理参数）。
  - 排序与筛选：按文件数、平均文件大小等指标快速定位高风险对象。

2.3.5.5 5) 任务视角：用列表定位慢任务并一键诊断

任务子域通常包含：

- **任务列表**：支持按任务名称/ID 搜索、时间范围快捷切换、多维筛选（用户/队列/类型/状态等）。
- **自动刷新**：适合追踪“正在跑”的任务；也支持手动刷新。
- **详情抽屉**：查看单个任务的状态、耗时、资源峰值等概览信息，并提供“诊断”入口联动到对话。

使用建议：

- 先用时间范围切到“最近 1h/6h/24h”，再用状态筛选聚焦在 Running/Failed 等目标集合。

2.3.5.6 6) 队列视角：从队列概览到资源洞察

队列子域常见子模块：

- **队列概览**：列表化展示队列与核心指标，适合做第一步排查。
- **队列分析 / 资源洞察（如有）**：用于观察队列结构分布、资源使用异常点。

2.3.5.7 7) PaaS 视角：云服务实例检索与详情

当工作空间接入云服务实例后，PaaS 子域常见能力包括：

- 按服务类型切换（如 Kafka/MySQL/Redis 等），并支持“全部”聚合视图。
- 模糊搜索：按实例名称/ID/地址/标签等搜索目标实例。
- 排序：按带宽/磁盘/Topic 等维度排序，结合升序/降序定位“异常头部实例”。
- 实例详情抽屉：点击实例卡片打开，查看资源列表与更多指标。

2.3.6 最佳实践

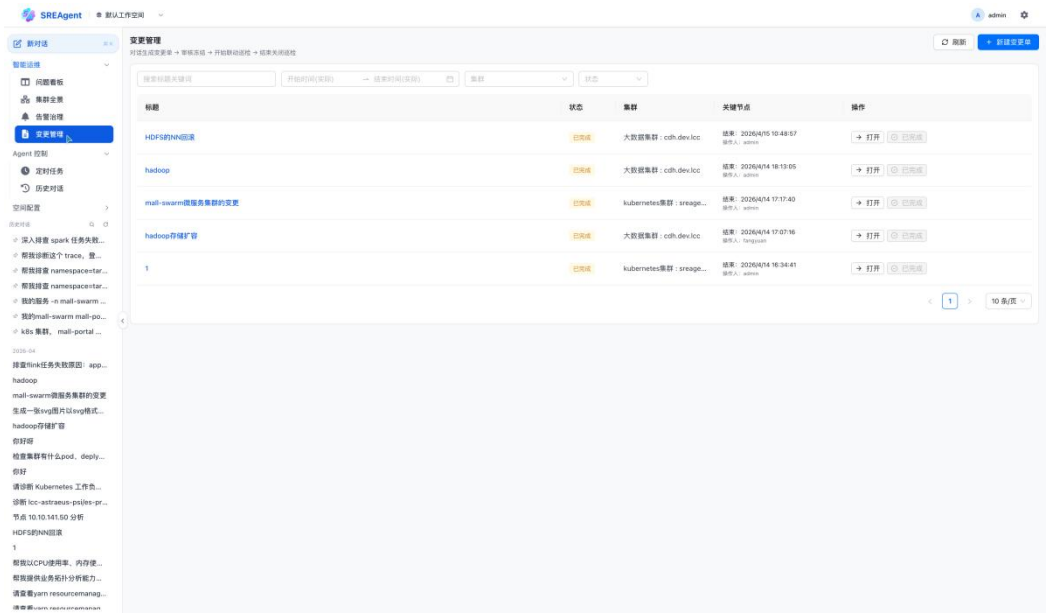
- **从全景到诊断**：在“任务/节点/实例”定位到异常对象后，优先使用页面提供的“诊断”入口进入对话，把上下文一次性带进去。
- **形成巡检习惯**：对高风险指标（小文件、资源峰值、失败率）建立固定的巡检路径，并沉淀为定时任务或 Runbook。
- **分享链接复现视角**：将带有当前视角的页面链接分享给同事，减少“你看到的和我看到的不一样”的沟通成本。

2.4 变更管理（Runbooks）

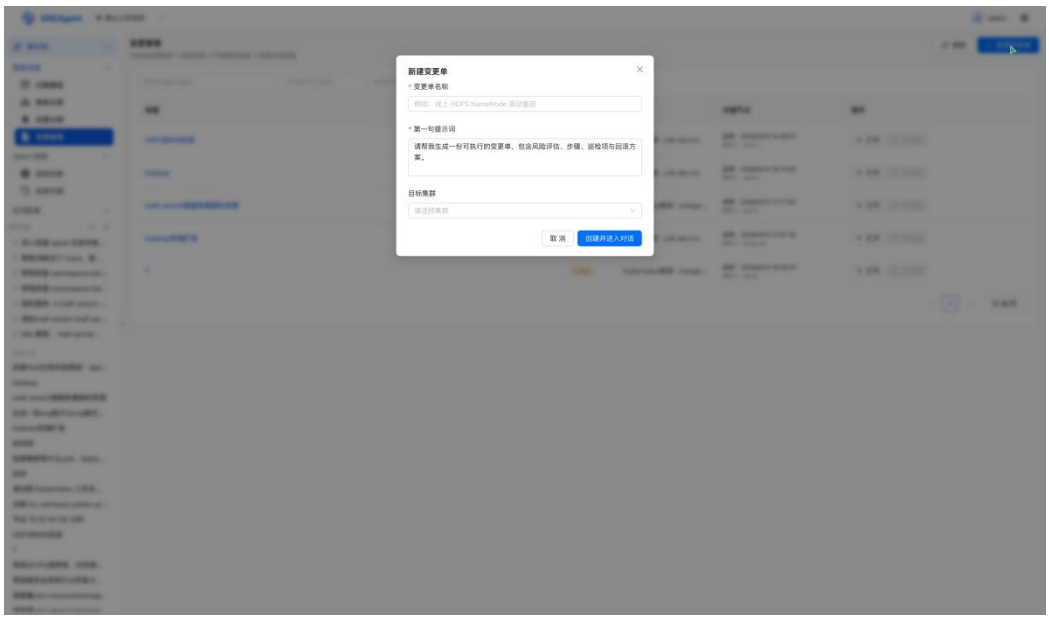
变更管理用于把“变更计划、执行步骤、风险控制、验证与复盘”沉淀成可审计的变更单（Runbook），并与对话诊断、定时巡检联动，实现“先写清楚再执行、执行中有护栏、执行后可追溯”的变更闭环。

2.4.1 页面概览（截图）

变更管理列表页：



新建变更单弹窗：



2.4.2 入口

- **菜单入口**：侧边栏「变更管理」。

2.4.3 核心概念

- **变更单（Runbook）**：一份可执行的变更文档，通常包含变更背景、执行步骤、回滚预案、验证口径、影响范围等内容（以实际模板为准）。
- **对话绑定**：每个变更单会绑定一个对话。进入变更单后，你既可以编辑文档，也可以直接在对话中追问细节，把结论沉淀进文档。
- **冻结与审计**：变更单在审核通过后进入冻结态，核心内容不可随意修改，用于保证执行过程中的一致性与可追溯。
- **变更护栏（联动巡检）**：开始变更后，系统会为该变更创建巡检任务，按固定间隔执行检查，持续监控执行过程中的风险信号。

2.4.4 生命周期（状态流转）

变更单通常遵循以下状态（以界面为准）：

- **草稿**：可编辑，可补充内容与证据。
- **审核中**：已提交审核，等待审核人处理。
- **已审核（冻结）**：审核通过，内容冻结，进入执行准备态。
- **执行中**：变更正在实施，巡检护栏任务生效。
- **已完成**：变更结束，护栏巡检关闭（保留记录用于复盘）。

常见动作（以界面按钮为准）：

- 提交审核、通过、驳回
- 开始变更、结束变更
- 查看巡检（跳转到定时任务的详情页）

2.4.5 核心操作指南

2.4.5.1 1) 新建变更单：把“首句提示词”当作自动生成模板的入口

在列表页点击“新建变更单”，通常需要填写：

- **变更单名称**：建议包含“对象 + 目标 + 窗口期”，例如“生产 Kafka 扩容（4/20 凌晨）”。
- **第一句提示词**：用于引导系统生成初稿或按你的要求补齐结构。

- **目标集群（可选）**：建议绑定变更目标集群，便于后续对话与巡检护栏复用该范围。

创建后系统会自动打开与该变更单绑定的对话，并将首句提示词作为对话起点。

2.4.5.2 2) 编辑与沉淀：在对话里问清楚，在 Runbook 里写清楚

推荐写作结构（可按团队模板调整）：

- 变更目标与范围
- 风险评估与影响面
- 前置检查（执行前必须满足的条件）
- 执行步骤（逐步可操作、可验证）
- 回滚预案（触发条件 + 回滚步骤）
- 验证口径（指标/日志/功能点）
- 复盘要点（执行结果、异常与改进项）

使用建议：

- 对话用于“把不确定性问清楚”，Runbook 用于“把确定的结论固化”。
- 把关键证据（日志片段、指标结论、影响对象清单）以可复制的形式写进文档，减少执行时临场查找成本。

2.4.5.3 3) 提交审核：把“冻结”当作执行口径的一部分

当变更单内容完整后提交审核：

- 审核通过后变更单进入冻结态，避免执行过程中不断改口径导致的风险。
- 若审核驳回，通常会回到草稿态，按审核意见补齐后重新提交。

2.4.5.4 4) 开始变更：启动“巡检护栏”并进入执行态

当变更单处于“已审核（冻结）”状态时，可点击“开始变更”：

- 系统会提示你配置巡检护栏（例如巡检名称、巡检内容、检查间隔、目标集群等，字段以弹窗为准）。
- 确认开始后，变更单进入“执行中”，并生成对应的巡检任务。
- 在执行过程中，可从变更单直接“查看巡检”，进入定时任务详情页查看巡检对话与结果。

2.4.5.5 5) 结束变更：关闭护栏并沉淀复盘

变更完成后点击“结束变更”：

- 系统会结束变更并关闭护栏巡检（通常为停用任务但保留审计记录）。
- 建议在变更单中补齐执行结果、异常与处理方式，并在复盘中沉淀为后续标准 SOP。

2.4.6 列表页的常用能力

变更管理列表页通常支持：

- **筛选/搜索**：按标题关键词、集群、状态、开始时间范围过滤。
- **关键节点审计**：展示创建/提交/通过/驳回/开始/结束等节点信息，便于快速复盘与追责。
- **行内操作**：不同状态下提供不同动作（提交审核/通过/驳回/开始/结束/查看巡检/打开对话等）。

2.4.7 最佳实践

- **先模板再细化**：用“第一句提示词”先生成结构化初稿，再逐项补齐证据与口径。
- **执行前冻结**：把审核通过后的冻结态当作流程保障，减少执行时的口径漂移。
- **护栏要可验证**：巡检内容优先写“可量化/可判定”的检查项（例如错误率、延迟、队列积压、关键日志关键字）。
- **变更后复用**：把成功的变更单沉淀为可复用模板，降低后续同类变更成本。

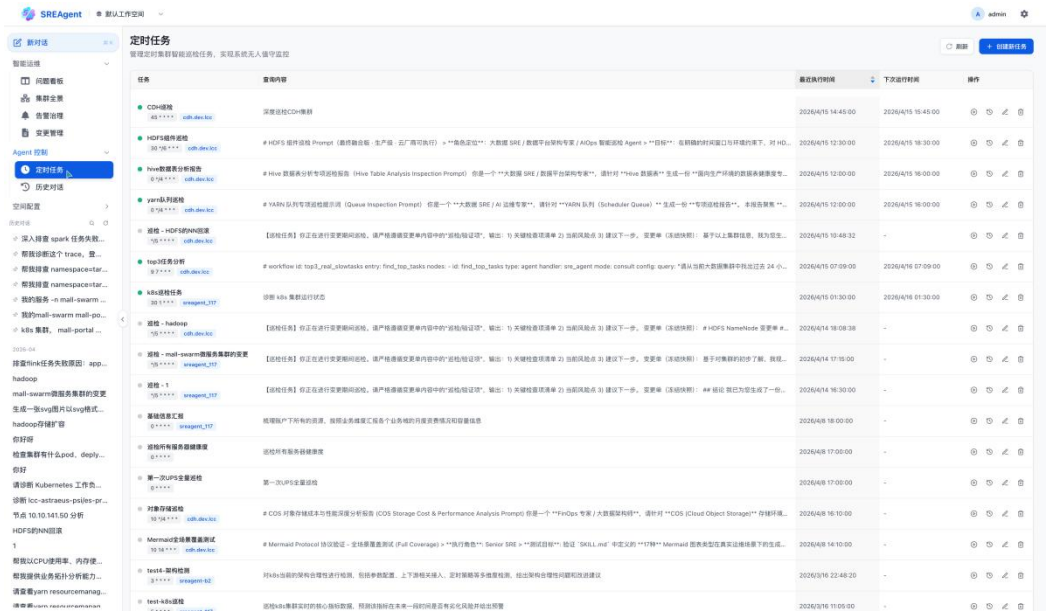
3 Agent 控制

3.1 定时任务（自动巡检）

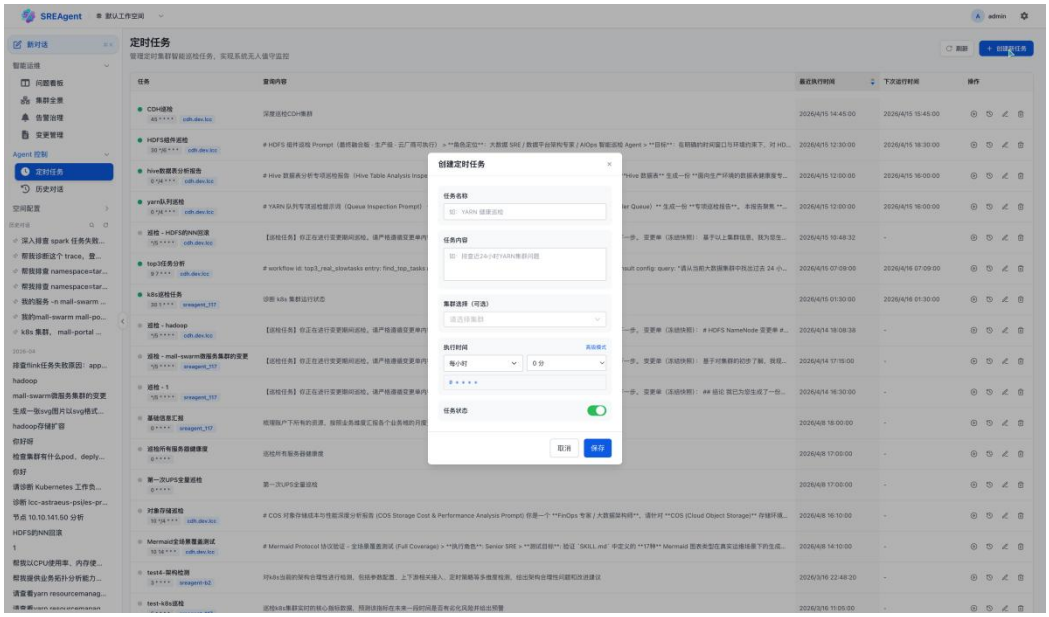
定时任务用于把“重复性体检工作”交给 Agent 定期执行：按计划触发对话型诊断，沉淀执行记录与问题清单，帮助值班团队实现“无人值守巡检 + 有问题自动沉淀 + 可复盘可追溯”。

3.1.1 页面概览（截图）

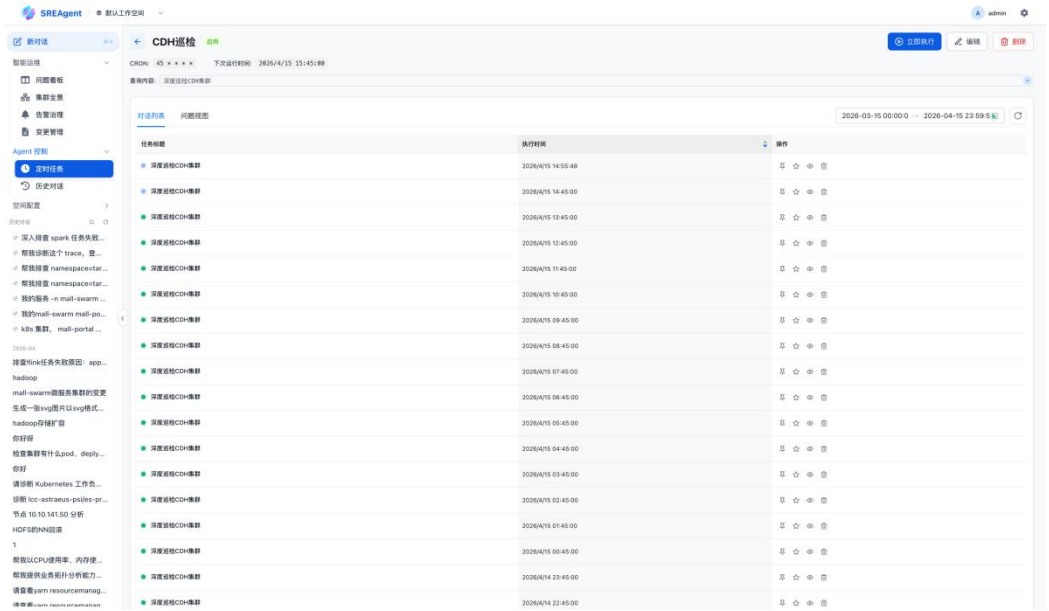
任务列表页：



创建任务弹窗：



任务详情页（执行记录）：



3.1.2 入口

- **菜单入口**：侧边栏「Agent 控制」->「定时任务」。

3.1.3 核心概念

- **巡检任务**：包含“任务名称 + 任务内容（自然语言）+ 执行时间（计划）+ 集群范围（可选）+ 启用状态”。

- **执行记录**：每次触发会生成一条对话记录（可从任务详情打开），用于复盘巡检的过程与结论。
- **问题沉淀**：部分版本会在任务详情中提供“问题视图”，将历次巡检对话中识别出的高频问题做聚合展示，便于跟进治理。

3.1.4 任务列表页

列表页用于统一管理工作空间内的巡检任务，常见能力包括：

- **刷新**：重新拉取任务列表。
- **创建新任务**：打开创建弹窗。
- **行内操作**：立即执行、查看历史（进入详情）、编辑、删除（通常带二次确认）。
- **排序与分页**：支持按“最近执行时间”等字段排序与分页浏览。

3.1.5 创建/编辑任务

点击“创建新任务”进入弹窗，按以下字段配置：

3.1.5.1 1) 任务名称

用于在列表与告警/复盘中快速识别任务目标，建议包含“范围 + 目标 + 频率”，例如：

- “生产集群 K8s 节点磁盘巡检（每 30 分钟）”
- “CDH 集群服务健康巡检（每天 08:00）”

3.1.5.2 2) 任务内容（Query）

任务内容是“每次巡检要做什么”的自然语言描述。推荐包含：

- **目标**：检查什么（节点、服务、队列、PVC、探针、告警、日志等）
- **时间窗口**：近 1h / 24h / 7d
- **输出规范**：结论摘要 + 异常清单 + 建议动作（最好要求按优先级排序）

3.1.5.3 3) 集群选择（可选）

若工作空间已配置多个集群，建议明确选择目标集群，避免巡检结果泛化或混淆。

3.1.5.4 4) 执行时间（计划）

支持以可视化方式配置常见执行频率，也支持高级模式直接填写 Cron 表达式。

常见示例：

- 0 8 * * *：每天 08:00 执行。
- */10 * * * *：每 10 分钟执行。

建议先低频运行，确认巡检内容有效后再提升频率，避免对数据源造成压力。

3.1.5.5 5) 任务状态

支持启用/停用。停用后任务不会自动触发，但可保留配置用于后续恢复。

3.1.6 执行与复盘

3.1.6.1 1) 立即执行

用于在创建或修改后验证任务内容与计划配置是否正确，也适合临时加测（例如上线前、变更后、告警频发时）。

3.1.6.2 2) 查看历史（任务详情）

任务详情页通常包含：

- **任务配置区**：任务名称、计划、下次运行时间、任务内容摘要等。
- **对话列表**：按时间范围筛选历次执行记录，支持刷新、排序、分页，并可对执行记录进行置顶/收藏等管理操作。
- **问题视图（如有）**：按任务聚合问题，展示问题标题、严重性/优先级、最近出现时间与建议动作，便于形成治理清单。

3.1.7 最佳实践

- **输出规范**：建议要求 Agent 输出“结论摘要 + 异常清单 + 建议动作”，并按优先级排序，便于值班快速同步。
- **任务分层**：将“高频轻量巡检”与“低频深度巡检”分层，避免任务过重导致执行时间过长。
- **复盘闭环**：对高频问题，结合“问题看板”沉淀为 Runbook/标准 SOP，并在工作空间治理项中固化执行口径。

3.2 新对话（智能对话）

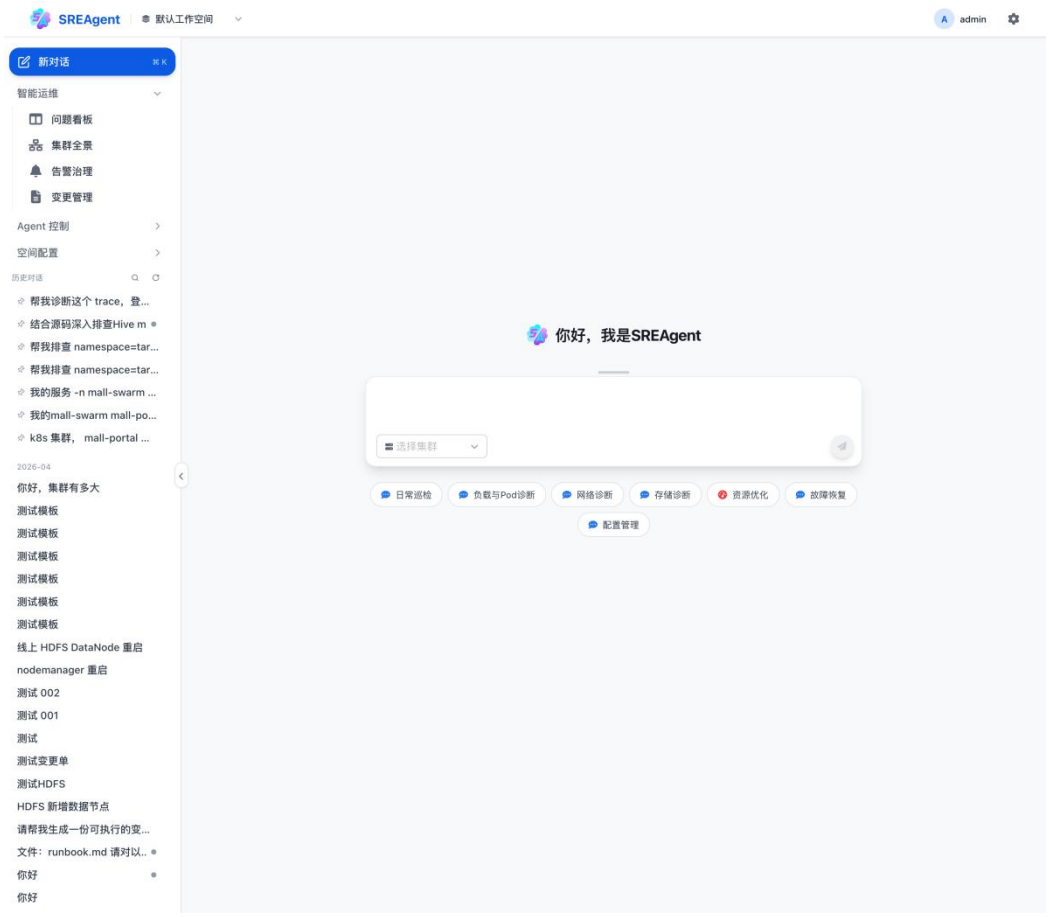
智能对话是 Agent 能力的核心入口，用于以自然语言发起诊断、分析与咨询。系统会以“对话 + 可追踪的执行过程”的方式工作：一边输出结论，一边展示关键步骤（计划、工具调用、执行结果），方便值班同学快速确认当前进度、证据链与下一步动作。

3.2.1 入口与快捷方式

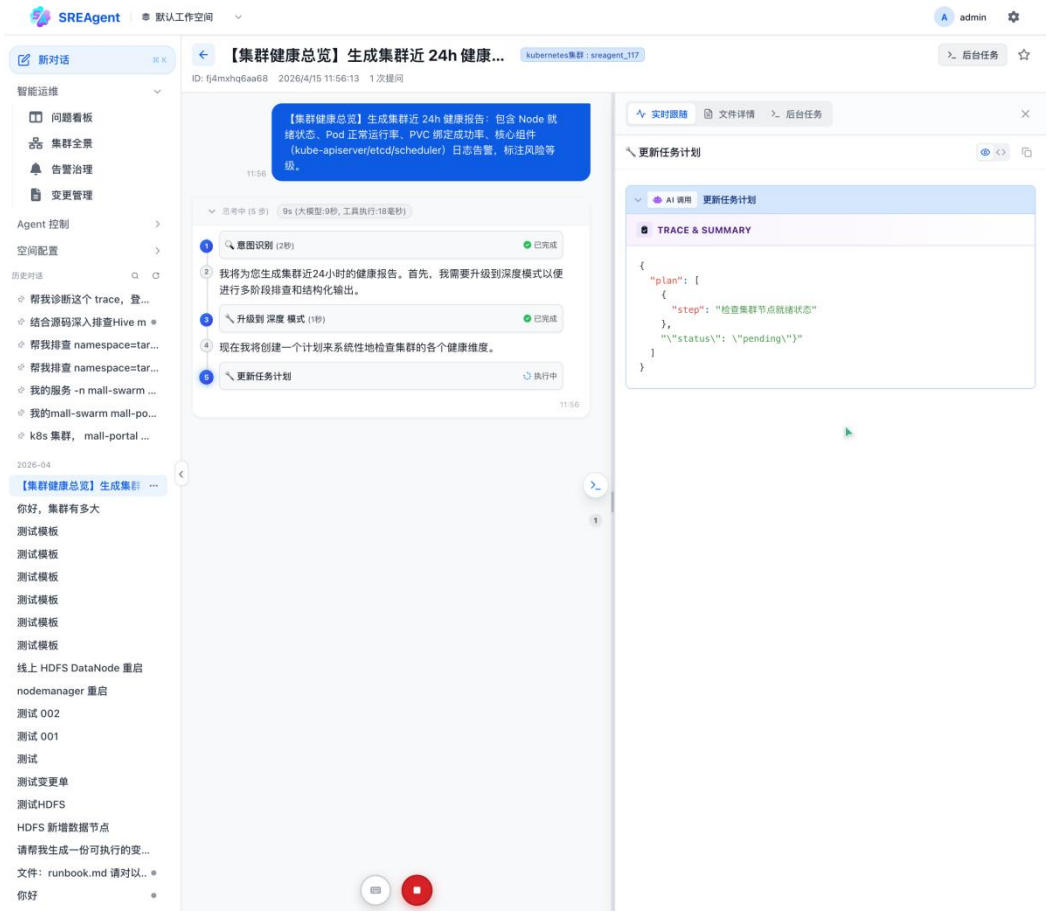
- **菜单入口**：侧边栏「新对话 ⌘ K」。
- **快捷键**：⌘ K 快速开启新对话。

3.2.2 页面概览（截图）

新建对话页（提示词模板 + 输入区 + 集群选择）：



对话详情页（消息流 + 右侧详情面板）：



3.2.3 新建对话

3.2.3.1 1) 工作空间与上下文

系统通常要求先选择工作空间后才能正常使用对话能力。在页面左上角可以切换当前所在的工作空间，不同工作空间的集群、告警源、任务与对话资产相互隔离。

3.2.3.2 2) 提示词模板（Prompt 模版）

新建对话页提供一组“场景化提示词模板”，用于快速发起高质量提问。常见分类例如：日常巡检、网络诊断、存储诊断、故障恢复等。

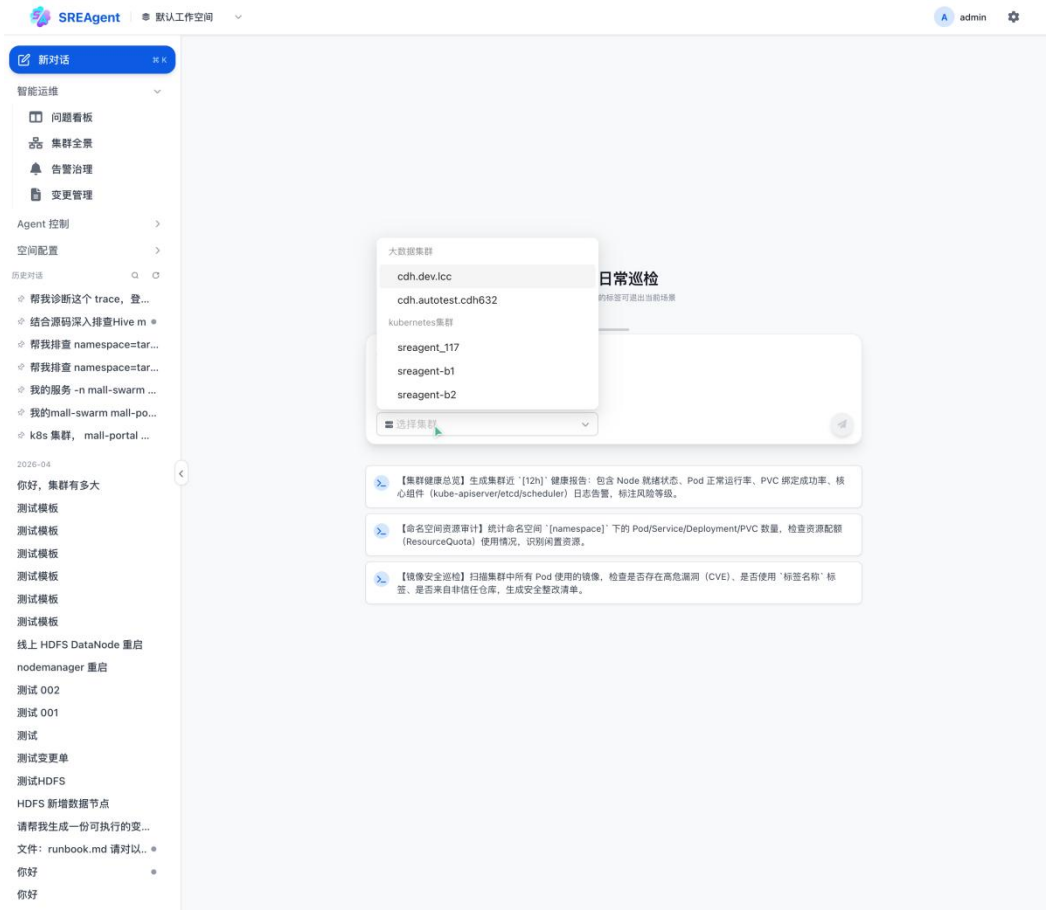
模板通常包含可编辑的占位符（例如时间范围、命名空间、资源名）。点击模板后，占位符会以“内联输入框”的形式出现在输入区，方便逐项替换。

使用建议：

- 先用模板把“范围/目标/输出格式”定下来，再补充日志片段、告警 ID、工单号等关键线索。
- 当占位符较多时，优先补全“时间范围 + 集群/命名空间/服务名”，通常能显著提升命中率。

3.2.3.3 3) 集群选择（Cluster）

对话支持绑定一个集群上下文，便于 Agent 在后续诊断中使用正确的数据源与配置。集群选择器支持搜索与下拉选择。



说明：

- 通常建议在发送前先选择集群，避免“无上下文”导致诊断结果过于泛化。
- 对话进入执行阶段后，集群可能会进入“锁定状态”，防止同一对话在中途切换上下文造成结果混乱。

3.2.3.4 4) 输入与发送

- **发送**：输入完成后按 Enter 发送。
- **换行**：使用修饰键（⌥/Shift/Ctrl/⌘）+ Enter 换行。
- **停止**：对话执行中可随时停止，适用于方向不对或信息不足的场景。

3.2.4 对话详情（过程透明与精细控制）

3.2.4.1 1) 消息流与消息类型

对话详情页的消息通常由两大部分构成：

- **用户消息**：每一次提问（包括运行中追加的提问）都会形成一条独立的问题节点，便于复盘与定位。

- **AI 步骤流**：Agent 的回复不仅包含自然语言结论，还会把执行过程拆解成可读的步骤流，常见类型包括：
  - **思考/解释**：为何要做这一步、判断依据是什么。
  - **计划（Plan）**：将排查拆成多步任务，并标注当前执行到哪一步。
  - **任务（Task）**：一次工具调用/一次分析动作，通常会携带状态（执行中、已完成、失败）。
  - **结果（Result）**：工具输出的摘要、报告片段、列表或文件型产出预览。

使用建议：

- 需要快速汇报时，优先抓“计划 + 当前步骤 + 风险结论”。
- 需要深挖证据链时，重点查看任务与结果卡片的细节内容。

3.2.4.2 2) 迷你状态（Mini）

对话执行中，输入区支持“迷你态”展示，避免遮挡消息内容，同时保留关键控制按钮：

- 一键展开输入框继续提问。
- 一键停止当前执行。
- 快速回到底部跟随最新输出。

3.2.4.3 3) 提问定位（问题导航）

当一条对话中包含多次提问时，可以使用右侧的问题导航能力：

- **跳转到第 N 个问题**：快速定位到某一次提问对应的上下文与回答。
- **问题预览**：鼠标悬停可快速确认每个问题的简要内容，避免翻页查找。

3.2.4.4 4) 运行中追加提问：排队 / 引导

当 Agent 正在执行时，仍然可以继续输入补充信息。系统提供两种追加模式：

- **排队（Queue）**：把追加内容放入队列，等待当前阶段结束后作为“下一条问题”正式进入对话流。适合补充关键线索（日志、告警 ID、时间范围）但不希望打断当前执行。
- **引导（Steer）**：把追加内容作为“对当前执行的即时纠偏/约束”。适合快速纠正方向，例如“只看某个命名空间”“优先排查网络层”。

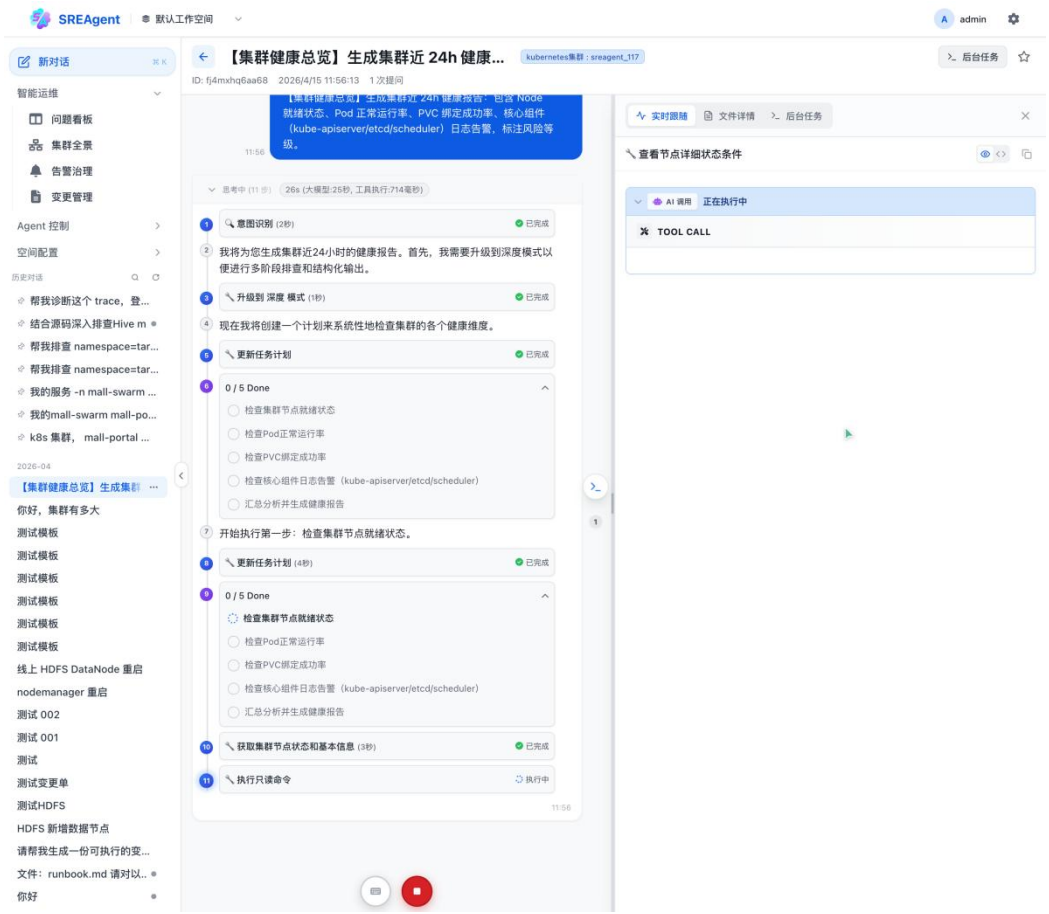
3.2.4.5 5) 右侧详情面板

对话详情页右侧提供可切换的详情面板，用于“跟随执行进度、打开步骤细节、管理后台任务”。

3.2.4.5.1 a) 实时跟随

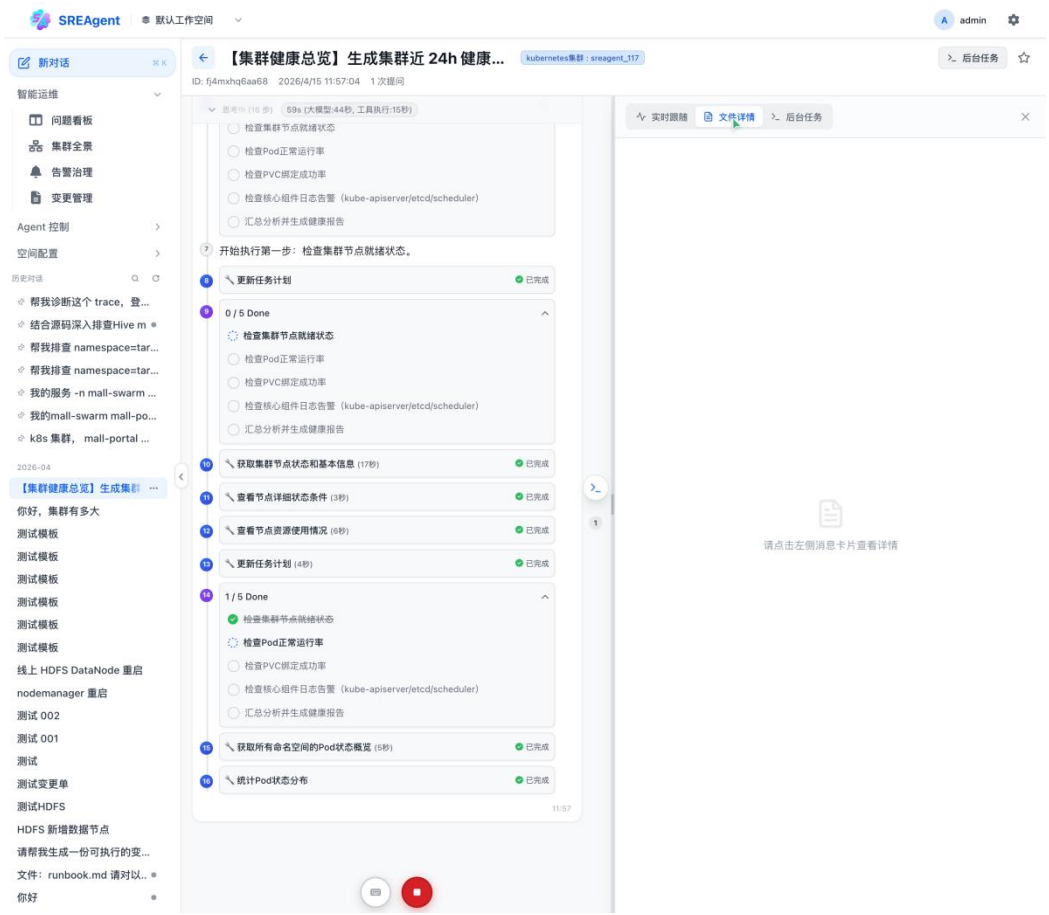
用于展示当前正在执行的关键任务，常见能力包括：

- 预览/源码视图切换。
- 一键复制内容，便于粘贴到工单或值班群同步。



3.2.4.5.2 b) 文件详情（任务/结果预览）

文件详情用于查看步骤卡片的细节：当你在左侧点击某个任务卡或结果卡时，右侧会展示对应内容（包括过程输出、报告片段、文件型结果预览等）。



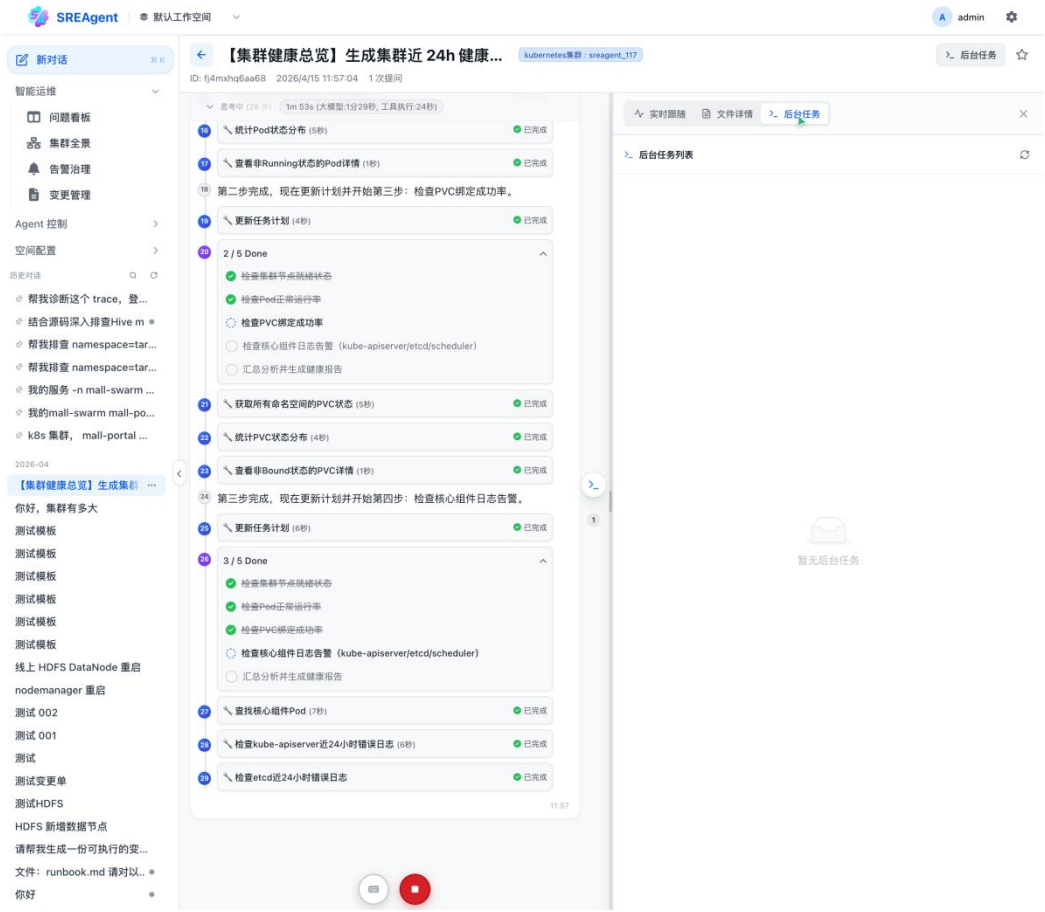
使用建议：

- 日常复盘时，优先在文件详情中提取“关键证据 + 关键结论 + 建议动作”。
- 若结果内容较长，可切换视图并复制关键段落用于沉淀 SOP。

3.2.4.5.3 c) 后台任务

后台任务用于查看对话过程中派发的异步执行任务（例如较耗时的查询、批量处理、生成报告等）。常见操作包括：

- 查看任务状态（排队中、执行中、已完成、失败、已取消）。
- 停止执行中的任务（必要时）。
- 打开输出详情，按需加载更多日志/结果片段。

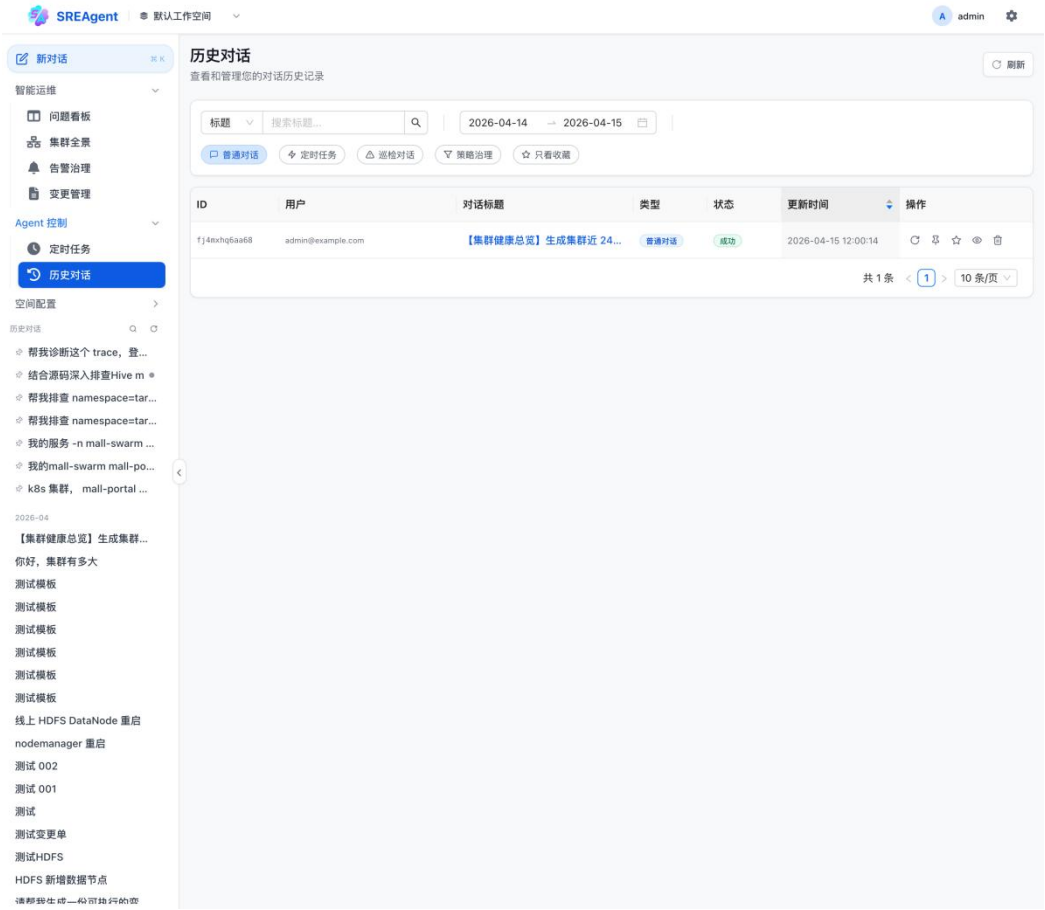

SREAgent
默认工作空间
admin
新对话
智能运维
问题看板
集群全景
告警治理
变更管理
Agent 控制
空间配置
历史对话
帮我诊断这个 trace，登...
结合源码深入排查Hive m
帮我排查 namespace=tar...
帮我排查 namespace=tar...
我的服务 -n mall-swarm ...
我的mall-swarm mall-po...
k8s 集群， mall-portal ...
2026-04
【集群健康总览】生成集群
你好，集群有多大
测试模板
测试模板
测试模板
测试模板
测试模板
测试模板
线上 HDFS DataNode 重启
nodemanager 重启
测试 002
测试 001
测试
测试变更单
测试HDFS
HDFS 新增数据节点
请帮我生成一份可执行的变...
文件：runbook.md 请对以...
你好
【集群健康总览】生成集群近 24h 健康...
kubernetes集群：sreagent_117
ID: fj4mxhq6aa68 2026/4/15 11:57:04 1 次提问
后台任务
思考中 (29 步) 1m 53s (大模型:1分29秒, 工具执行:24秒)
统计Pod状态分布 (5秒) 已完成
查看非Running状态的Pod详情 (1秒) 已完成
第二步完成，现在更新计划并开始第三步：检查PVC绑定成功率。
更新任务计划 (4秒) 已完成
2 / 5 Done
检查集群节点就绪状态
检查Pod正常运行率
检查PVC绑定成功率
检查核心组件日志告警（kube-apiserver/etcd/scheduler）
汇总分析并生成健康报告
获取所有命名空间的PVC状态 (5秒) 已完成
统计PVC状态分布 (4秒) 已完成
查看非Bound状态的PVC详情 (1秒) 已完成
第三步完成，现在更新计划并开始第四步：检查核心组件日志告警。
更新任务计划 (6秒) 已完成
3 / 5 Done
检查集群节点就绪状态
检查Pod正常运行率
检查PVC绑定成功率
检查核心组件日志告警（kube-apiserver/etcd/scheduler）
汇总分析并生成健康报告
查找核心组件Pod (7秒) 已完成
检查kube-apiserver近24小时错误日志 (6秒) 已完成
检查etcd近24小时错误日志 已完成
11:57
实时跟随
文件详情
后台任务
后台任务列表
暂无后台任务

3.3 历史对话

历史对话用于管理和检索工作空间内的历史会话，包括快速搜索、时间范围筛选、按类型过滤、收藏/置顶与删除等能力，帮助把一次排障过程沉淀为可复用资产。

3.3.1 页面概览（截图）

历史对话列表页（筛选区 + 列表 + 行内操作）：



3.3.2 入口与定位方式

- **菜单入口**：侧边栏「Agent 控制」->「历史对话」。
- **快捷入口**：侧边栏顶部的「搜索历史」可直接进入历史对话，并聚焦到搜索框。

3.3.3 侧边栏“历史对话”列表（快速入口）

侧边栏会展示近期的对话记录，用于快速回到最近看过的对话，适合在值班过程中频繁来回切换。

常见能力：

- **分组展示**：置顶对话会单独置顶显示，其余对话通常按月份分组。
- **状态提示**：可从列表项的状态提示快速判断对话是否仍在运行、失败或已结束。
- **更多操作**：列表项通常支持“重试、置顶/取消置顶、收藏/取消收藏、删除”等快捷操作。

说明：

- 侧边栏列表更偏“快捷入口”，完整筛选与管理建议使用“历史对话”列表页。
- 若界面出现“重命名功能暂未开放”的提示，表示当前版本暂不支持从侧边栏直接提交重命名。

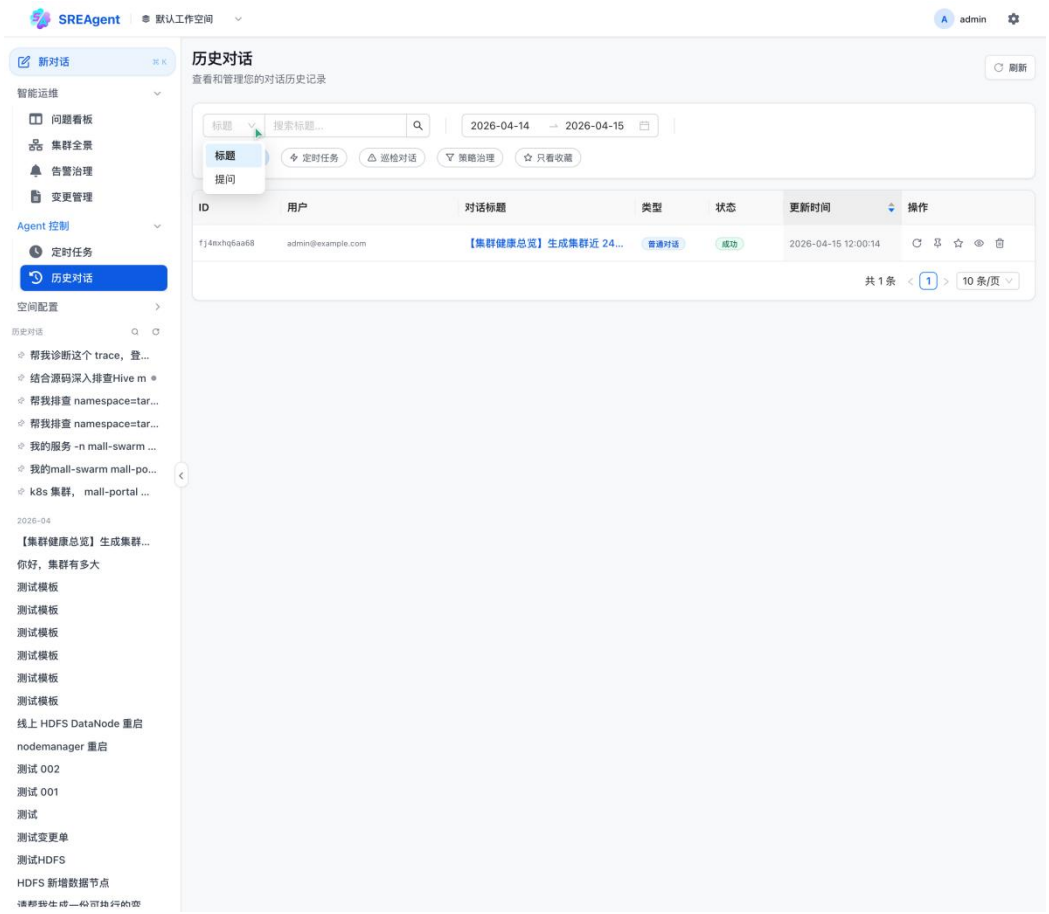
3.3.4 历史对话列表页（/history）

3.3.4.1 1) 搜索与筛选

3.3.4.1.1 a) 搜索范围：标题 / 提问

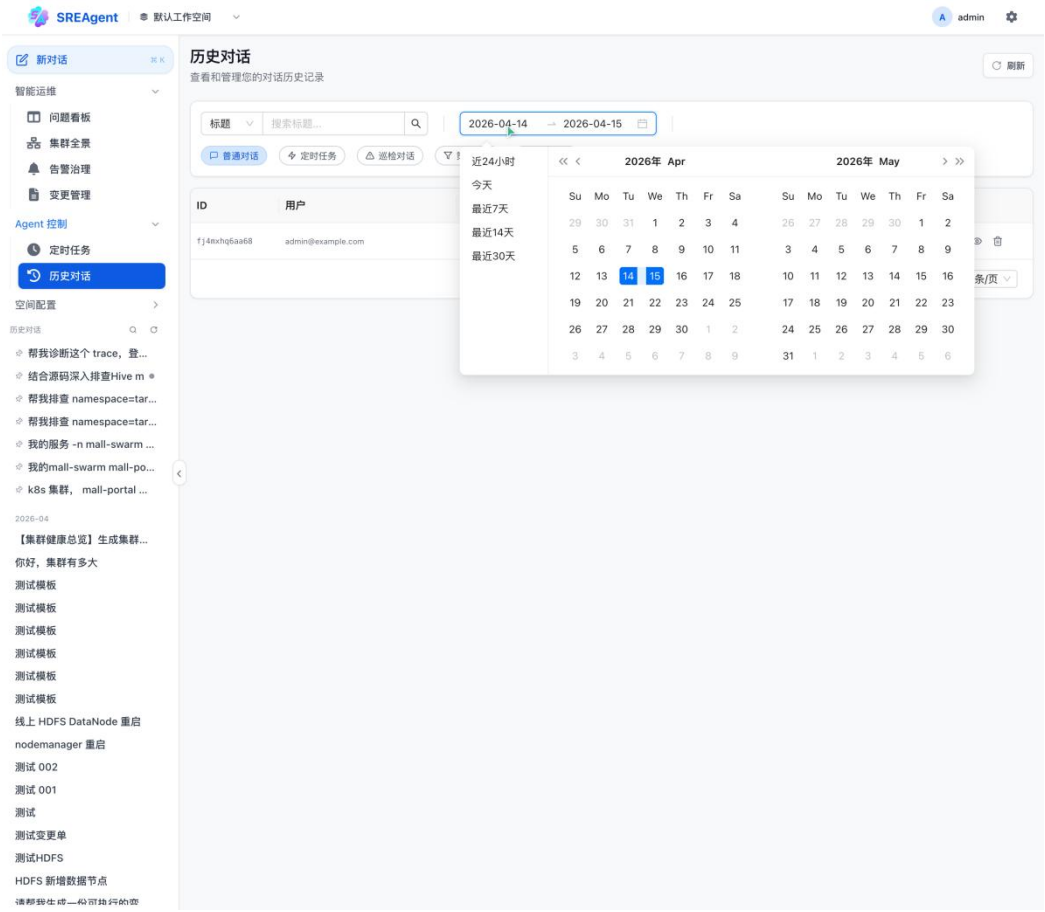
搜索支持切换范围：

- **标题**：按对话标题匹配。
- **提问**：按用户提问内容匹配（更适合“记得问过什么，但忘了标题”的场景）。



3.3.4.1.2 b) 时间范围：快速预设 + 自定义

时间范围支持快速预设（例如“近 24 小时、今天、最近 7 天、最近 14 天、最近 30 天”）以及手动选择起止日期。



使用建议：

- 值班排障复盘优先使用“近 24 小时/最近 7 天”，减少噪声。
- 若对话列表为空，先检查时间范围是否选得过窄。

3.3.4.1.3 c) 对话类型与只看收藏

列表页支持按对话类型进行过滤（例如普通对话、定时任务、巡检对话、策略治理等），也支持“只看收藏”快速聚焦沉淀案例。

3.3.4.2 2) 列表信息与排序

列表常见字段包括：

- **ID**：对话标识，用于排查与定位。
- **用户**：发起人（常用于审计与协作）。
- **标题**：可点击进入对话详情。
- **类型**：以标签形式展示，不同类型的跳转与能力略有差异。
- **状态**：运行中、成功、失败、已取消等。
- **更新时间**：支持按更新时间排序，便于快速找到最近变化的对话。

3.3.4.3 3) 行内操作（对话资产管理）

对每条对话通常支持以下操作（以界面为准）：

- **重试**：从该对话复用上下文重新发起一次诊断，适合“当时信息不足/策略变更后再跑一次”。
- **置顶/取消置顶**：将高频使用的标准案例置顶，提升可见性。
- **收藏/取消收藏**：沉淀团队可复用的故障案例与 SOP。
- **查看详情**：进入对话详情页复盘完整过程。
- **删除**：清理无效或测试会话（删除前通常会有二次确认）。

说明：

- 若类型为“定时任务”，在列表中可能支持跳转到对应的任务详情页面，便于追溯该对话的触发来源。

3.3.5 进入对话详情（复盘与证据链）

点击历史对话标题或“查看详情”即可进入对话详情页，复盘对话上下文、步骤流与最终结论。

复盘建议：

- 先确认“问题、范围、时间窗口”是否清晰，再核对关键步骤与关键证据。

- 需要沉淀 SOP 时，建议把“结论摘要 + 证据链 + 建议动作”提取到 Runbook 或工单系统中。

3.3.6 最佳实践

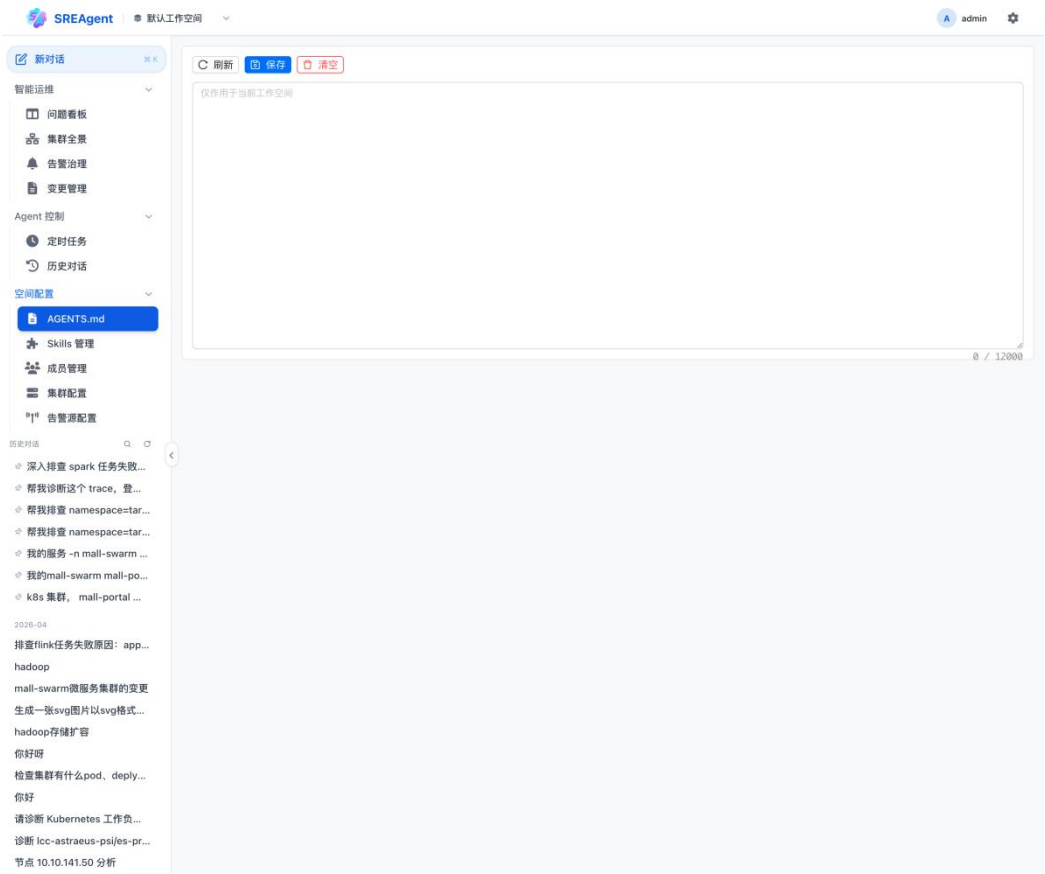
- **标准命名**：建议标题包含“系统/组件 + 现象 + 日期”，例如“k8s 控制面抖动 2025-11-24”。
- **收藏策略**：将“高复用案例 / 高风险故障案例 / 关键 SOP”收藏，便于团队新同学学习。
- **置顶策略**：将“本周/本月高频问题”“核心系统故障案例”置顶，形成值班同学的快速入口。
- **复盘闭环**：将最终结论与证据链整理到 Runbook 或工单系统中，形成完整的处理闭环。

4 空间配置

4.1 AGENTS.md（工作空间注入）

AGENTS.md 用于为当前工作空间配置一段“对话注入内容”：它会作为团队的统一背景与约束，帮助 Agent 更贴近你的真实环境与工作方式，减少“反复解释上下文”的成本。

4.1.1 页面概览（截图）



4.1.2 入口

- **菜单入口**：侧边栏「空间配置」->「AGENTS.md」。

4.1.3 适用场景

- **统一输出口径**：例如“先结论后证据链”“必须给回滚方案”“风险分级标准”等。
- **沉淀环境背景**：集群命名、核心业务链路、常见故障模式、关键指标口径等。
- **写清楚边界与合规**：敏感信息处理规则、不可访问的数据源、账号权限说明等。

提示：AGENTS.md 只作用于当前工作空间，不影响其他工作空间。

4.1.4 权限说明

- **工作空间管理员**：可编辑并执行“保存/清空”。
- **普通成员**：只读查看（按钮不可用或不展示）。

4.1.5 核心操作指南

4.1.5.1 1) 刷新

当你怀疑内容被其他管理员修改，或希望获取最新版本时，点击“刷新”重新加载内容。

4.1.5.2 2) 保存（管理员）

编辑完成后点击“保存”生效。

使用建议：

- 保存前先在内容中写清楚“当前工作空间/当前环境”的范围，避免跨环境复用造成误导。
- 将“固定口径”放前面，“可变信息”放后面，便于后续维护。

4.1.5.3 3) 清空（管理员）

当需要临时解除注入约束，或准备重建内容时，可点击“清空”。清空通常会弹出二次确认，避免误操作。

4.1.6 推荐模板（可直接复制）

```
# 工作空间背景

– 环境：生产 / 预发 / 测试

– 核心业务：订单、支付、风控
```

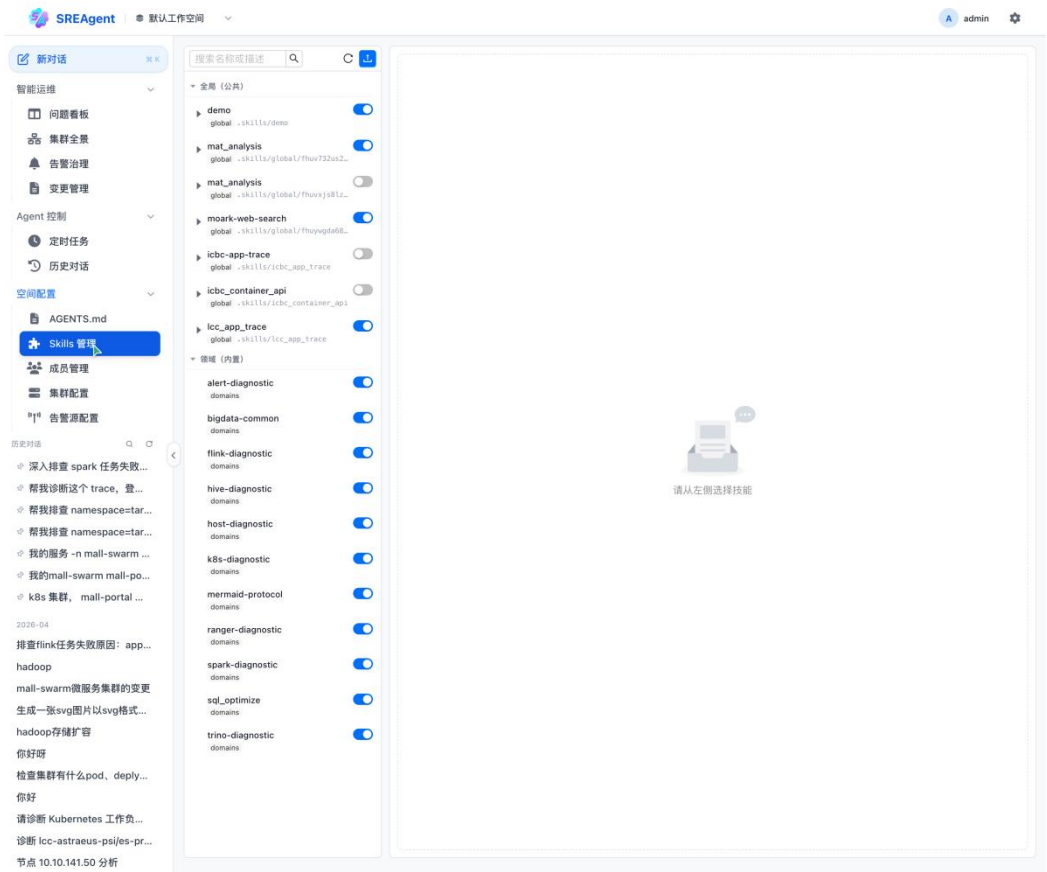
```
– 关键链路：下单 -> 支付 -> 履约

# 集群与命名规范
– K8s：prod-k8s-01 / prod-k8s-02
– 大数据：prod-emr-01
– 告警对象命名：{服务名}-{实例}-{机房}

# 输出格式要求（强约束）
1. 先给结论（可执行动作）
2. 再给证据链（日志/指标/事件，按时间顺序）
3. 最后给建议动作（按优先级排序：止血 -> 恢复 -> 根治）
4. 每条建议动作必须包含：风险、回滚方案、验证手段

# 值班规则（示例）
– P0：业务不可用，必须 5 分钟内响应
– P1：核心链路降级，必须 15 分钟内响应

# 合规与安全
– 禁止输出任何密钥/Token/密码
– 涉及敏感数据只做脱敏展示（保留前后 2 位）
```

4.2 Skills 管理

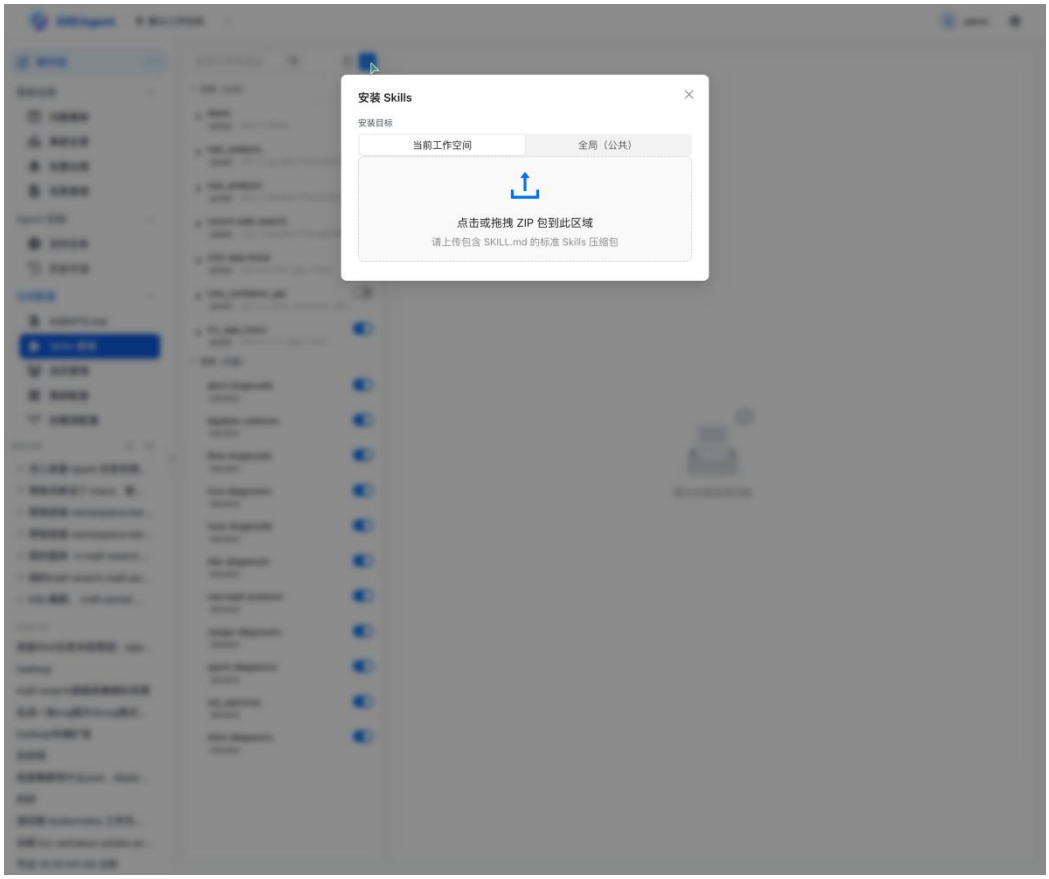
Skills 是 Agent 可调用的“工具包”。通过 Skills 管理，你可以把团队常用的排障脚本、诊断流程、数据查询能力封装成标准能力，并在工作空间内持续治理（安装、启停、配置、更新、下线）。

4.2.1 页面概览（截图）

Skills 管理页（左侧列表 + 右侧详情/编辑区）：



安装 Skills 弹窗（上传 ZIP + 选择作用域）：



4.2.2 入口

- **菜单入口**：侧边栏「空间配置」->「Skills 管理」。

4.2.3 核心概念

4.2.3.1 1) Skills 的作用域

平台通常会把 Skills 分为三类（以界面展示为准）：

- **工作空间 Skills**：仅当前工作空间可见与可用，适合沉淀“团队私有”的排障工具与 SOP。
- **全局（公共）Skills**：对所有工作空间可见，适合沉淀“公司级”通用能力。
- **领域（内置）Skills**：平台内置的领域能力，通常仅支持启用/禁用，不支持在线编辑与查看详情（以按钮状态为准）。

4.2.3.2 2) “启用/禁用”是治理的核心开关

Skills 的存在与否通常由“是否安装/是否在系统中发现”决定；是否可被 Agent 真实使用，还取决于：

- 是否处于启用状态
- 是否满足平台的运行限制（例如某些技能仅在特定模式下可用）

实践建议：把“启用”当作上线，把“禁用”当作熔断。出现误触发、能力重叠或风险脚本时，优先禁用再排查。

4.2.3.3 3) Skills 包的基本形态（面向上传/安装）

上传安装通常要求 ZIP 包包含必要的技能说明与脚本目录（以系统校验提示为准）。建议在团队内统一技能包模板，避免“能装但不能跑”的情况。

4.2.4 权限说明

- **工作空间管理员**：可管理工作空间 Skills（安装/更新/删除/配置/文件编辑等）。
- **系统管理员**：可管理全局（公共）Skills（通常也具备工作空间 Skills 的管理能力）。
- **普通成员**：通常只读查看，并使用已启用 Skills（不可执行管理操作，以按钮是否可点击为准）。

4.2.5 页面结构

- **左侧**：Skills 列表（按作用域分组）+ 搜索 + 刷新 + 安装入口 + 文件树。
- **右侧**：当前选中 Skill 的详情（说明文档）、配置编辑器、文件编辑器、测试运行入口等。

4.2.6 核心操作指南

4.2.6.1 1) 搜索与刷新

- **搜索**：按名称或描述搜索技能，快速定位目标能力。
- **刷新**：用于让页面立刻发现新安装/新更新的技能；当你通过“上传/更新/下线”修改后，建议刷新确认状态一致。

4.2.6.2 2) 安装 Skills（ZIP）

点击“安装 Skills”后：

- 选择安装目标：**当前工作空间** 或 **全局（公共）**（是否可选取决于你的权限）。
- 上传 ZIP 包完成安装。

使用建议：

- 在上线前先安装到工作空间验证，再考虑升级为全局（公共）能力。
- 上传前确保包内包含完整依赖与说明文档，避免“安装成功但运行失败”。

4.2.6.3 3) 启用/禁用 Skills

启用开关通常在左侧列表与右侧详情页都可操作：

- **启用**：表示允许 Agent 使用该技能。
- **禁用**：立即停止该技能对 Agent 的可见性（常用于应急熔断）。

使用建议：

- 对“高风险操作类技能”（例如会改配置、会执行写操作）建议默认禁用，仅在需要时启用并设置明确的使用边界。
- 当技能能力与现有技能高度重叠时，建议只启用一个，避免 Agent 选择混乱。

4.2.6.4 4) 查看说明文档（Readme）

选中某个 Skill 后，右侧默认展示该技能的说明文档（Markdown 渲染）。建议把以下信息写在说明里：

- 适用场景与输入要求（需要哪些参数/上下文）
- 输出结构约束（结论/证据/建议动作）
- 风险提示与回滚说明
- 示例（最小可用示例 + 典型示例）

4.2.6.5 5) 配置 Skills（JSON 配置）

右侧“配置”入口通常提供 JSON 编辑器：

- 用于填写运行参数、外部连接信息引用（例如 API Key 的引用方式）、沙箱资源限制等（以系统支持项为准）。
- 支持快捷保存（例如 Cmd/Ctrl + S）。

安全建议：

- 不要把明文密钥写入配置文件内容；优先使用平台提供的安全存储/引用机制（以实际部署为准）。

4.2.6.6 6) 文件管理与在线编辑（文件树）

当你展开技能文件树后，通常可以：

- **新建文件/目录**：用于补齐脚本、Prompt、辅助模块等。
- **删除文件/目录**：用于清理废弃实现（注意：删除往往不可恢复）。
- **在线编辑并保存**：支持快捷保存（例如 Cmd/Ctrl + S）。

使用建议：

- 把脚本与文档保持同仓治理：修改脚本后同步更新说明文档与示例。
- 对生产空间，优先通过“更新技能包”做版本化发布，避免直接在线改动造成不可追溯。

4.2.6.7 7) 测试运行（在线调试）

部分技能会提供“测试运行/在线调试”入口：

- 支持输入参数并执行，查看运行输出（常见为 Markdown 报告或结构化结果）。
- 适合在上线前快速验证依赖、参数与输出格式。

4.2.6.8 8) 下载 / 更新 / 删除

右侧详情通常提供：

- **下载**：导出 ZIP 便于备份/迁移。
- **更新**：上传 ZIP 全量覆盖（通常有风险提示）；建议作为“版本发布”使用。
- **删除**：卸载技能（高风险）；建议先禁用观察，再删除。

4.2.7 常见问题（FAQ）

4.2.7.1 Q1：为什么有些技能不能打开详情，只能开关启用?

可能是“领域（内置）Skills”，通常仅允许启用/禁用，不支持查看详情或编辑（以界面限制为准）。

4.2.7.2 Q2：为什么提示缺少依赖（例如检测到 requirements 但缺少 libs）?

很多部署场景下运行环境不支持动态安装依赖。建议采用以下策略之一（以团队约定为准）：

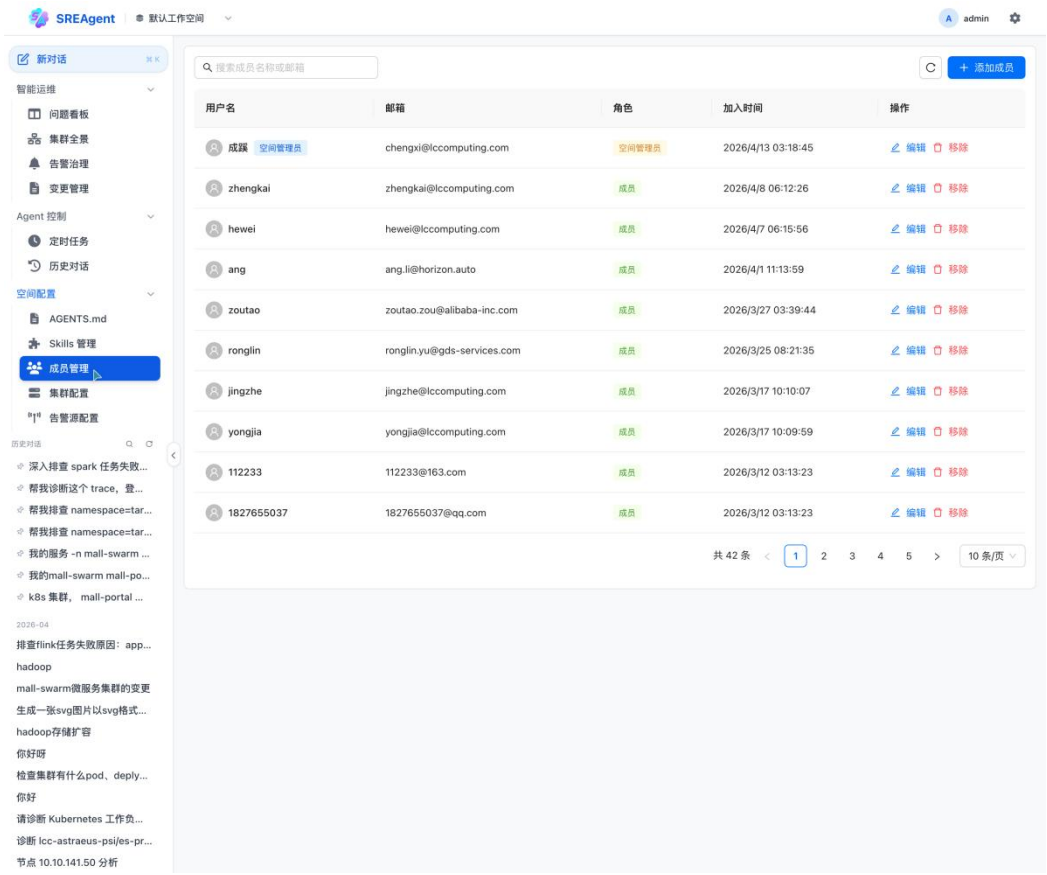
- **自包含依赖**：将第三方依赖打包到技能目录的 libs/ 中再上传。
- **指定运行镜像**：在技能配置中指定预装依赖的运行镜像（如果部署支持）。

4.3 成员管理

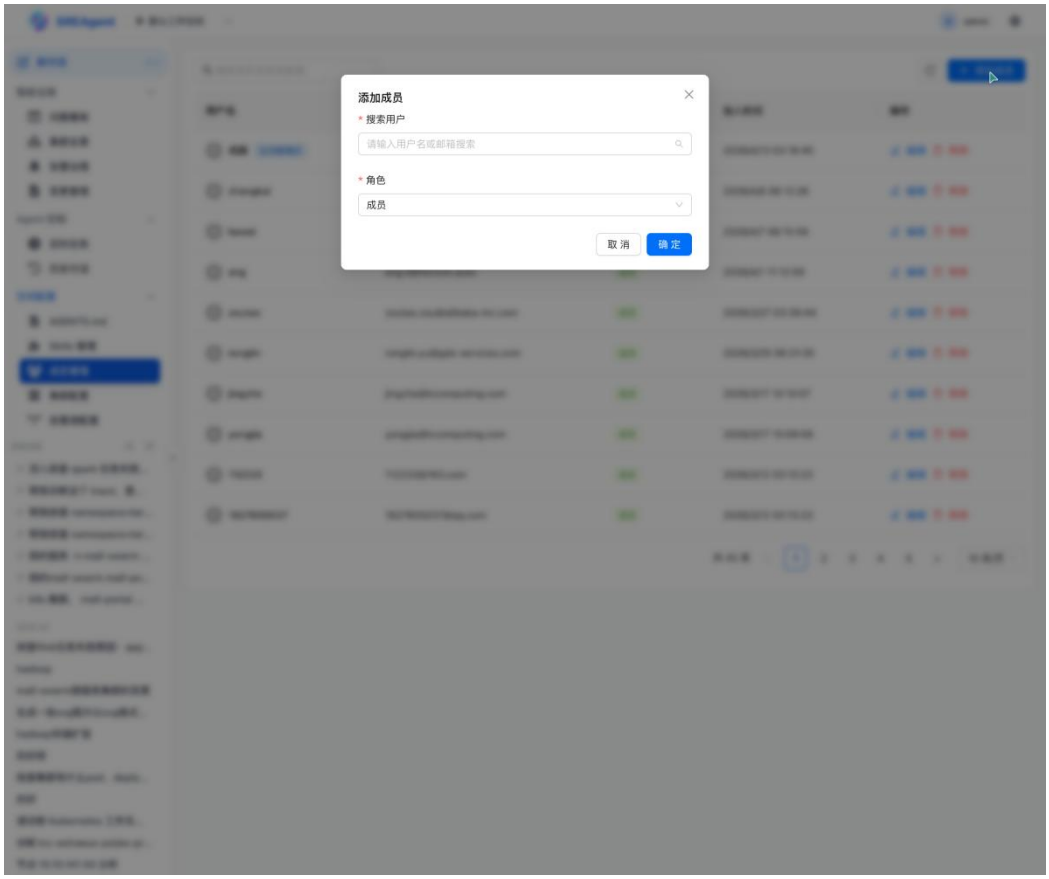
成员管理用于维护当前工作空间的成员列表与角色。角色会直接影响“空间配置”中的可操作范围（例如 Skills 管理、集群配置、告警源配置等），建议把它当作空间治理的第一入口。

4.3.1 页面概览（截图）

成员列表页：



添加成员弹窗：



4.3.2 入口

- **菜单入口**：侧边栏「空间配置」->「成员管理」。

4.3.3 权限说明

- **工作空间管理员**：可添加成员、调整成员角色、移除成员。
- **普通成员**：只读查看成员列表。

4.3.4 页面字段说明

成员列表通常包含（以界面为准）：

- **用户名**：常带头像展示。
- **邮箱**：用于检索与身份确认。
- **角色**：空间管理员 / 成员。
- **加入时间**：用于审计与复盘。

4.3.5 核心操作指南

4.3.5.1 1) 搜索成员

在搜索框输入“用户名或邮箱关键字”，即可快速定位目标成员。

使用建议：

- 交接班/人员变动时，优先确认“空间管理员”名单是否符合现状，避免出现无人维护或权限过宽。

4.3.5.2 2) 添加成员（管理员）

操作步骤：

5. 点击“添加成员”。
6. 在“搜索用户”中输入用户名/邮箱关键字，选择目标用户。
7. 选择角色并确认。

角色选择建议：

- **空间管理员**：建议仅授予少数负责治理的同学，用于维护集群/告警源/Skills 等关键配置。
- **成员**：适合日常值班与排障使用，通常只需要查看与使用权限。

4.3.5.3 3) 调整成员角色（管理员）

在成员行点击“编辑”，选择新角色并保存。

使用建议：

- 将“临时管理员”明确在团队流程中留痕，变更完成后及时降权，减少长期风险。

4.3.5.4 4) 移除成员（管理员）

在成员行点击“移除”，二次确认后执行。

使用建议：

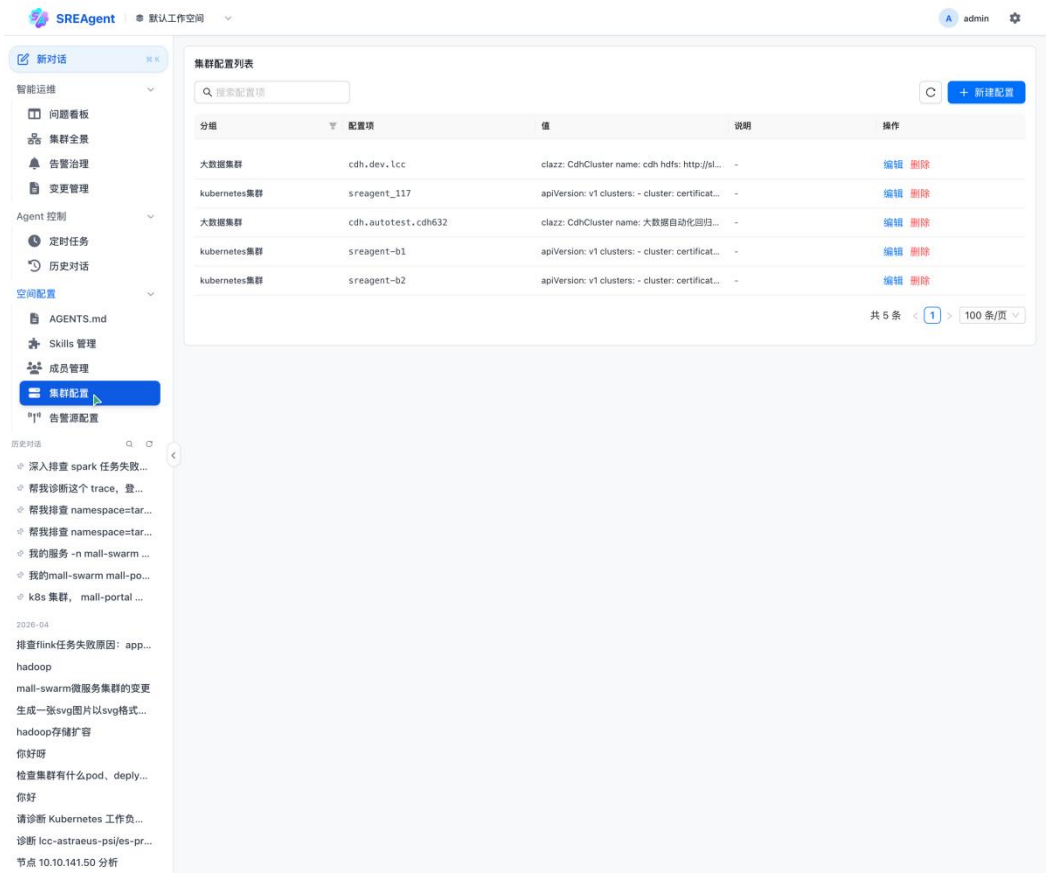
- 对已离组/离职成员及时移除，避免遗留权限。
- 对关键空间，建议配合“至少 2 名管理员”的治理原则，避免单点。

4.4 集群配置

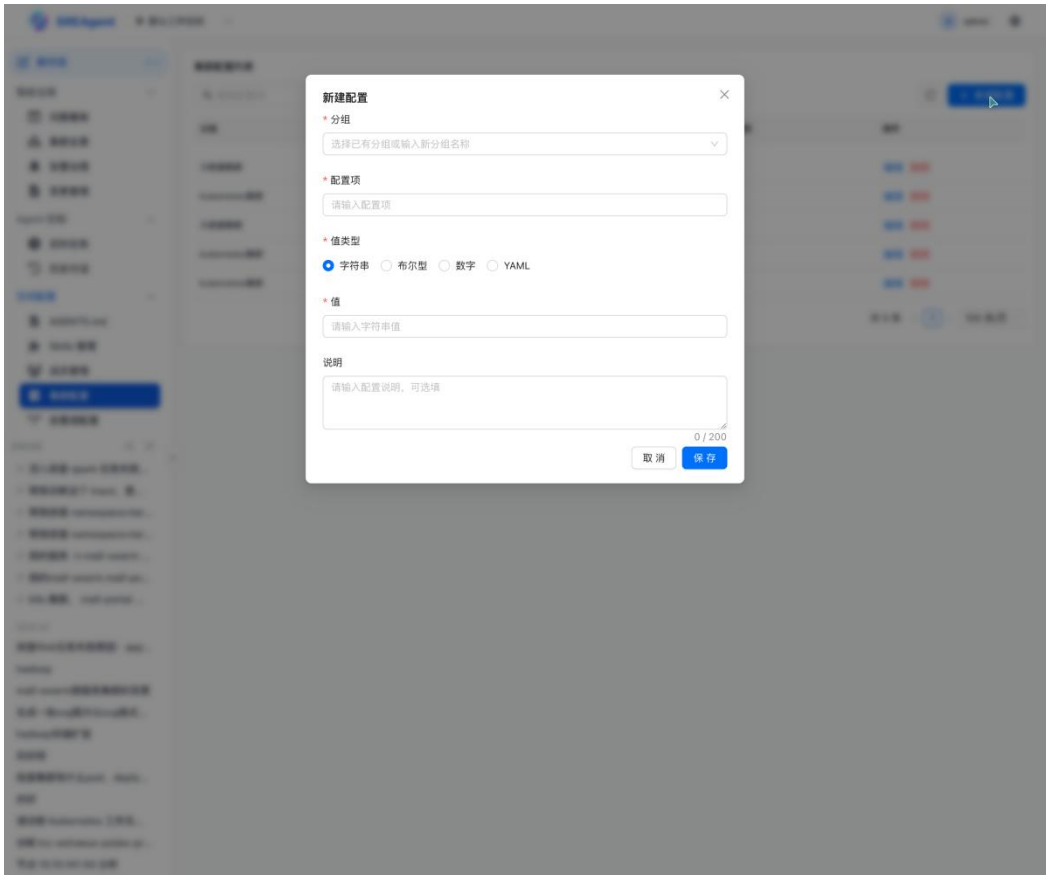
集群配置用于维护当前工作空间可用的“集群/环境清单”。它会被多个模块复用（例如对话诊断的集群选择、集群全景的视角切换、定时任务的执行范围绑定等），因此建议优先把这里配置完整。

4.4.1 页面概览（截图）

集群配置列表页：



新建配置弹窗：



4.4.2 入口

- **菜单入口**：侧边栏「空间配置」->「集群配置」。

4.4.3 权限说明

- **工作空间管理员**：可新建、编辑、删除。
- **普通成员**：只读查看。

4.4.4 配置项组织方式

集群配置采用“分组 + 配置项 + 值”的方式组织：

- **分组**：用于按环境/业务线拆分（例如 prod/staging/dev）。
- **配置项**：建议填写集群的唯一标识（便于在对话/全景/任务中直接定位）。
- **值类型与值**：支持字符串/布尔/数字/YAML，复杂配置优先用 YAML。
- **说明**：建议写清楚用途、负责人、接入方式，便于后续维护。

4.4.5 核心操作指南

4.4.5.1 1) 搜索与刷新

- **搜索**：在搜索框输入关键字，按配置项快速检索。
- **刷新**：当你刚修改了配置或怀疑数据不同步时，使用刷新重新加载列表。

4.4.5.2 2) 新建配置（管理员）

点击“新建配置”，按弹窗填写：

- **分组**：可选择已有分组，也可输入新分组。
- **配置项**：建议使用统一命名规范（例如 prod-k8s-01、prod-emr-01）。
- **值类型**：推荐 YAML 用于结构化信息（例如集群类型、访问地址、标签等）。
- **值**：按类型填写；YAML/JSON 类内容请确保格式正确。
- **说明**：建议包含“来源、责任人、变更记录口径”。

4.4.5.3 3) 编辑与删除（管理员）

- **编辑**：修改配置内容或说明，建议配合团队变更流程留痕。
- **删除**：高风险操作，会有二次确认；删除后相关模块下拉选项可能同步减少。

4.4.6 最佳实践

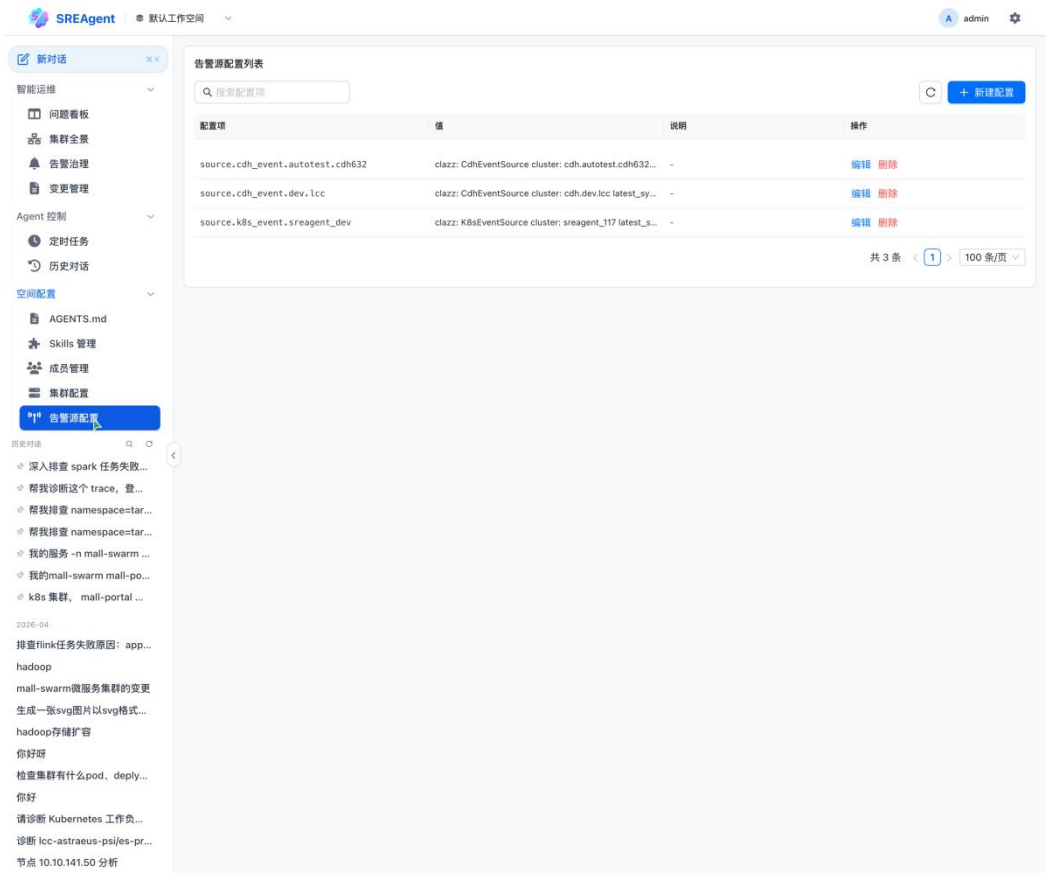
- **分组按环境拆分**：prod/staging/dev 是最常见且最稳定的分组方式。
- **配置项可读且稳定**：把“人类可读”和“可长期稳定”作为命名优先级，减少后续迁移成本。
- **配置即资产**：为关键集群补齐说明字段，明确责任人、用途和注意事项，避免“只有配置没有背景”的问题。

4.5 告警源配置

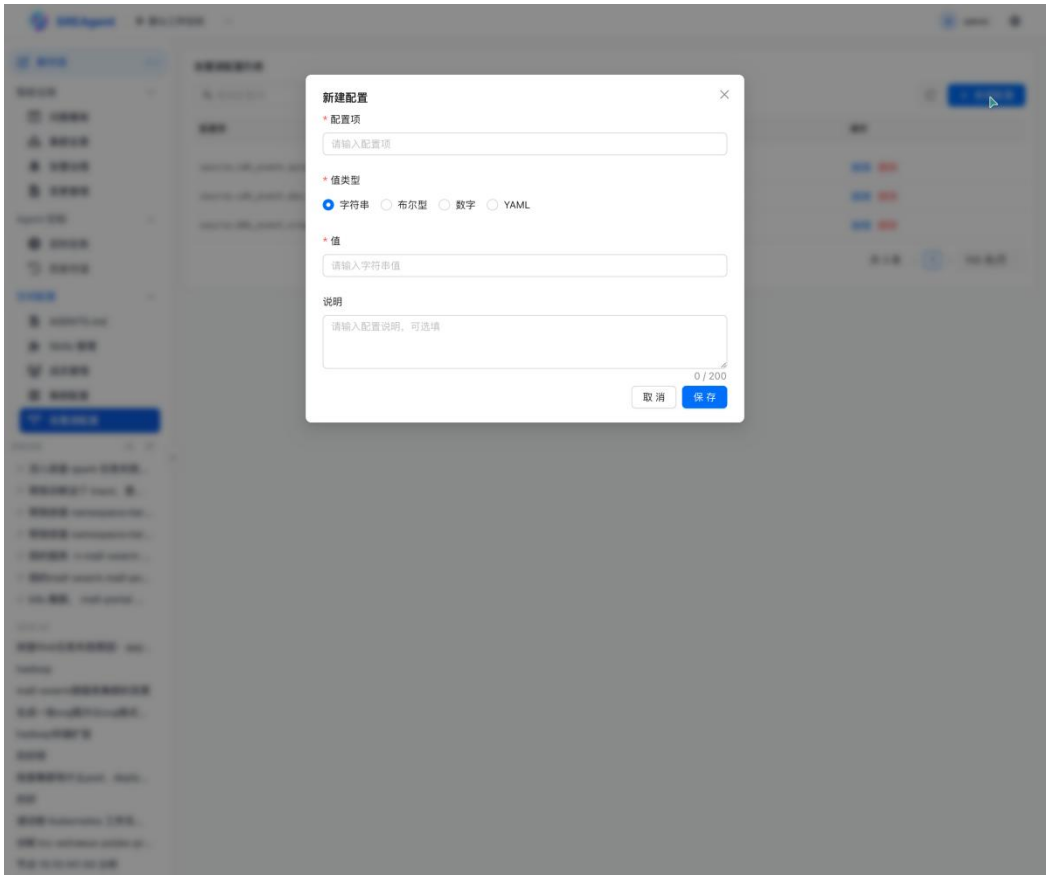
告警源配置用于维护当前工作空间的“告警接入通道”。告警治理模块能否正常展示告警列表、告警上下文是否完整，通常取决于这里的配置是否正确、是否与实际告警平台/数据源匹配。

4.5.1 页面概览（截图）

告警源配置列表页：



新建配置弹窗：



4.5.2 入口

- **菜单入口**：侧边栏「空间配置」->「告警源配置」。

4.5.3 权限说明

- **工作空间管理员**：可新建、编辑、删除。
- **普通成员**：只读查看。

4.5.4 配置项组织方式

告警源配置以“配置项 + 值”的方式组织（分组通常为系统固定，不需要手动选择）：

- **配置项**：用于区分不同告警源，或同一告警源的不同环境/租户/通道。
- **值类型与值**：支持字符串/布尔/数字/YAML；复杂接入建议用 YAML。
- **说明**：建议写清楚“来源系统、责任人、用途、回调地址/范围”等，便于排查问题。

4.5.5 核心操作指南

4.5.5.1 1) 搜索与刷新

- **搜索**：在搜索框输入关键字，按配置项快速检索。
- **刷新**：当告警平台侧配置调整后，建议刷新确认本页配置与预期一致。

4.5.5.2 2) 新建告警源配置（管理员）

点击“新建配置”，按弹窗填写：

- **配置项**：建议包含“来源类型 + 环境/租户”信息，例如 grafana_webhook_prod、prometheus_alertmanager_staging。
- **值类型**：推荐 YAML，便于表达结构化字段。
- **值**：通常包括告警源地址、鉴权方式引用、解析/映射规则等（以你的告警平台与接入协议为准）。
- **说明**：建议写清楚适用范围与维护人。

安全建议：

- 不要把明文密钥直接写进配置内容；优先使用平台提供的安全存储/引用方式（以实际部署为准）。

4.5.5.3 3) 编辑与删除（管理员）

- **编辑**：用于调整接入信息或解析规则；建议配合告警平台的变更记录。
- **删除**：高风险操作，会有二次确认；删除后告警治理的可用数据源可能减少或中断。

4.5.6 排障建议

当你在“告警治理”中遇到告警为空、字段缺失、解析异常时，优先检查：

- 本页是否存在对应的告警源配置项。
- 值（尤其是 YAML）是否格式正确。
- 是否误删/误改了关键通道，或责任人变更后未同步维护。

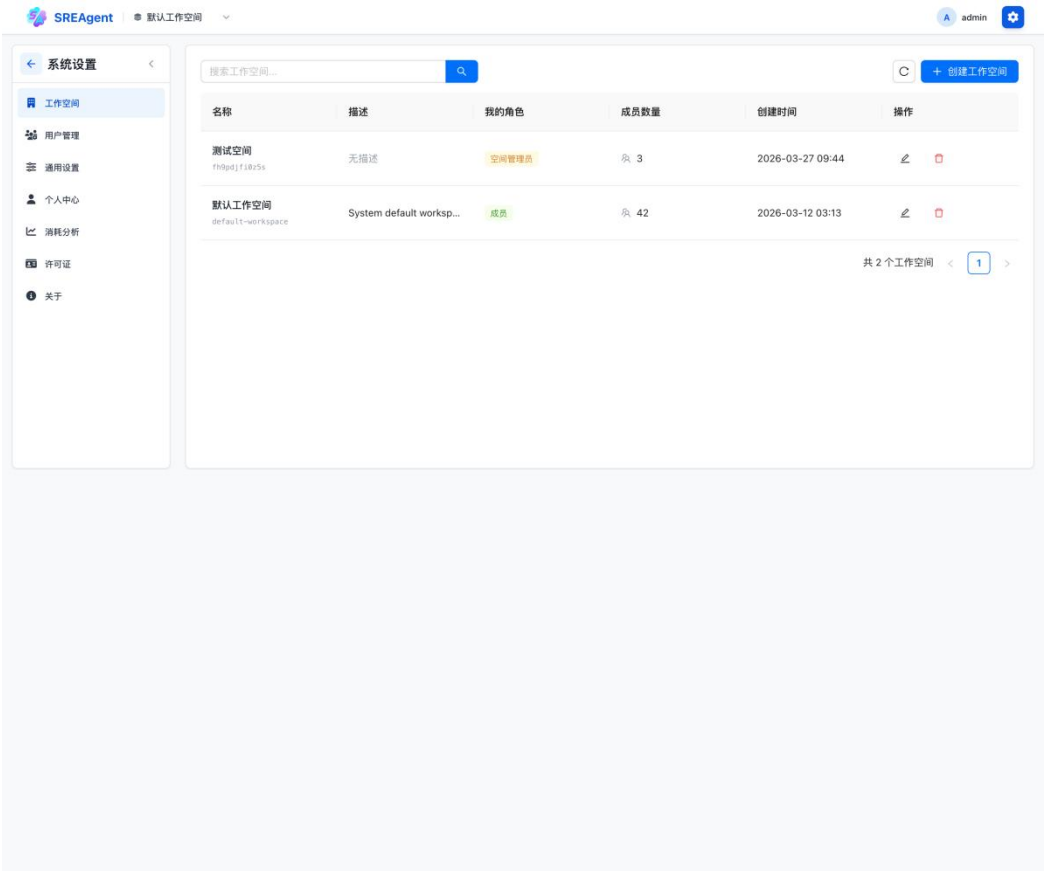
5 系统设置

5.1 工作空间管理（系统设置）

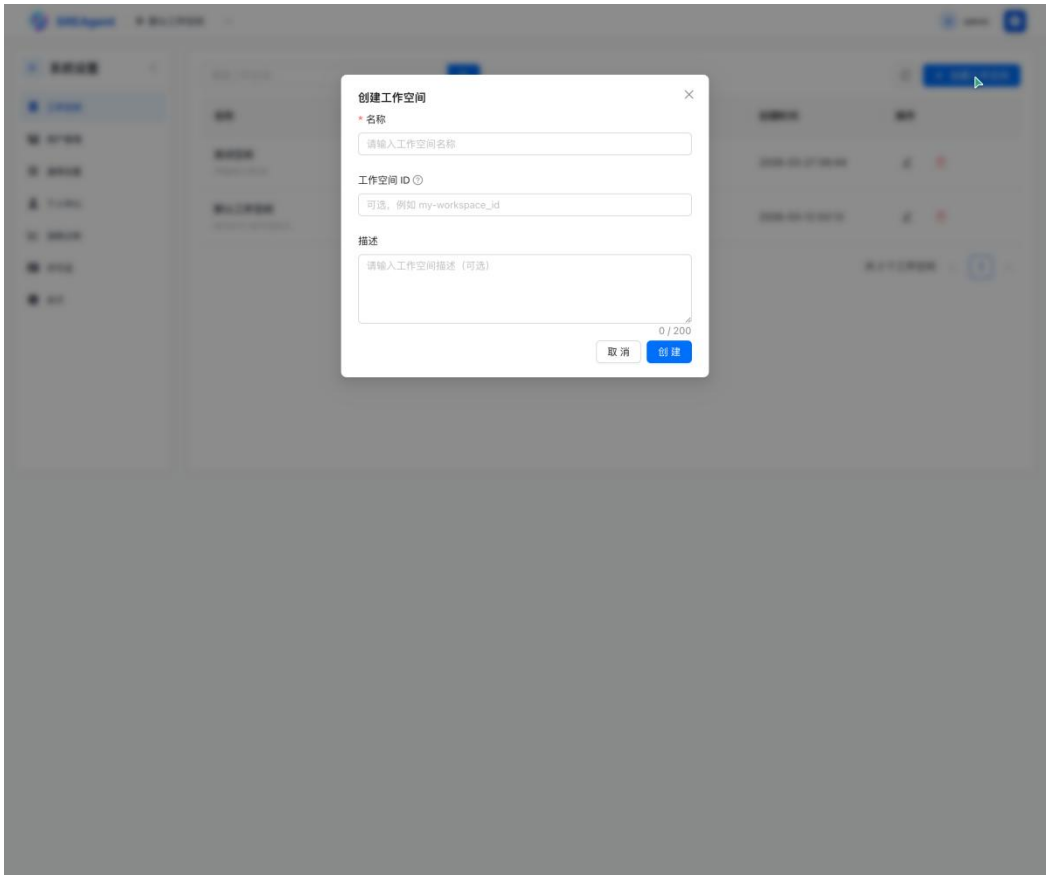
工作空间管理用于维护“工作空间生命周期”：创建新工作空间、编辑工作空间信息、删除无效工作空间，并查看自己在各工作空间的角色与成员规模。

5.1.1 页面概览（截图）

工作空间列表页：



创建工作空间弹窗：



5.1.2 入口

- **入口**：右上角齿轮进入「系统设置」，默认进入工作空间管理页。

5.1.3 权限说明

以按钮是否可见/可点击为准，常见规则：

- **系统管理员**：可创建、编辑、删除任意工作空间。
- **工作空间管理员**：通常只能编辑自己所在的工作空间（不可删除）。
- **普通成员**：只读查看列表。

5.1.4 页面字段说明

列表通常包含（以界面为准）：

- **名称**：工作空间展示名称。
- **工作空间 ID**：用于唯一标识（常用于 URL、权限与配置隔离）。
- **描述**：用于补充业务线/用途信息。
- **我的角色**：空间管理员 / 成员。
- **成员数量**：便于评估空间规模与治理成本。

- **创建时间**：用于审计与复盘。

5.1.5 核心操作指南

5.1.5.1 1) 搜索与刷新

- **搜索**：输入关键字快速定位工作空间。
- **刷新**：当你刚完成创建/编辑/删除后，可刷新确认列表已同步。

5.1.5.2 2) 创建工作空间（系统管理员）

点击“创建工作空间”，在弹窗中填写：

- **名称**：必填，建议用“业务线 + 环境/用途”命名。
- **工作空间 ID**：可选；不填时系统自动生成。建议只使用小写字母、数字与 `-` `/` `_`，并保持稳定，避免后续迁移成本。
- **描述**：可选；建议写清楚用途、负责人、关联系统范围。

5.1.5.3 3) 编辑工作空间信息

在列表行点击“编辑”，通常允许修改：

- 名称
- 描述

实践建议：

- 把“空间用途/边界/负责人”写在描述中，减少后续沟通成本。

5.1.5.4 4) 删除工作空间（高风险，系统管理员）

删除通常为不可恢复的高风险操作（以弹窗提示为准），建议遵循：

- 先确认该空间内的对话资产、定时任务、配置项是否已迁移或下线。
- 先在团队内完成审批/备案，再执行删除。

5.1.6 与“空间配置”的关系

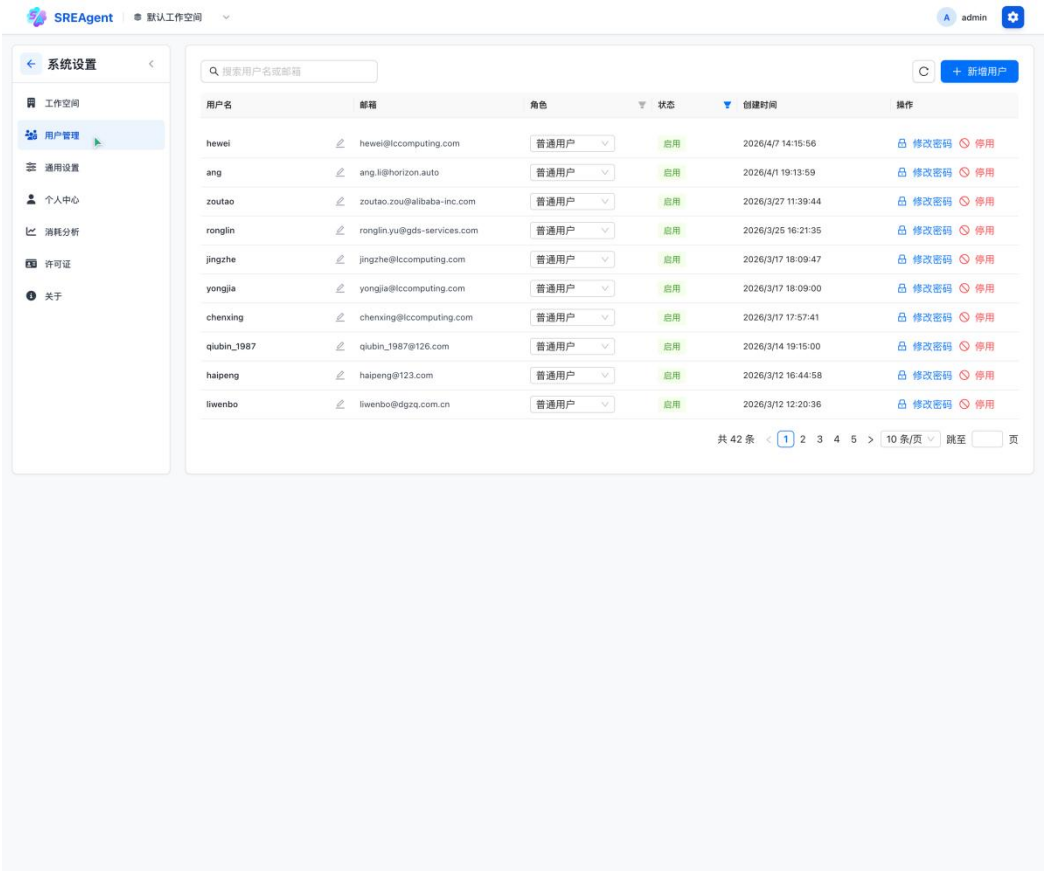
- **系统设置 -> 工作空间管理**：负责“空间的创建/编辑/删除”等全局治理动作。
- **空间配置（侧边栏）**：负责“空间内资产的日常配置”，例如成员管理、Skills 管理、集群配置、告警源配置等。

5.2 用户管理（系统设置）

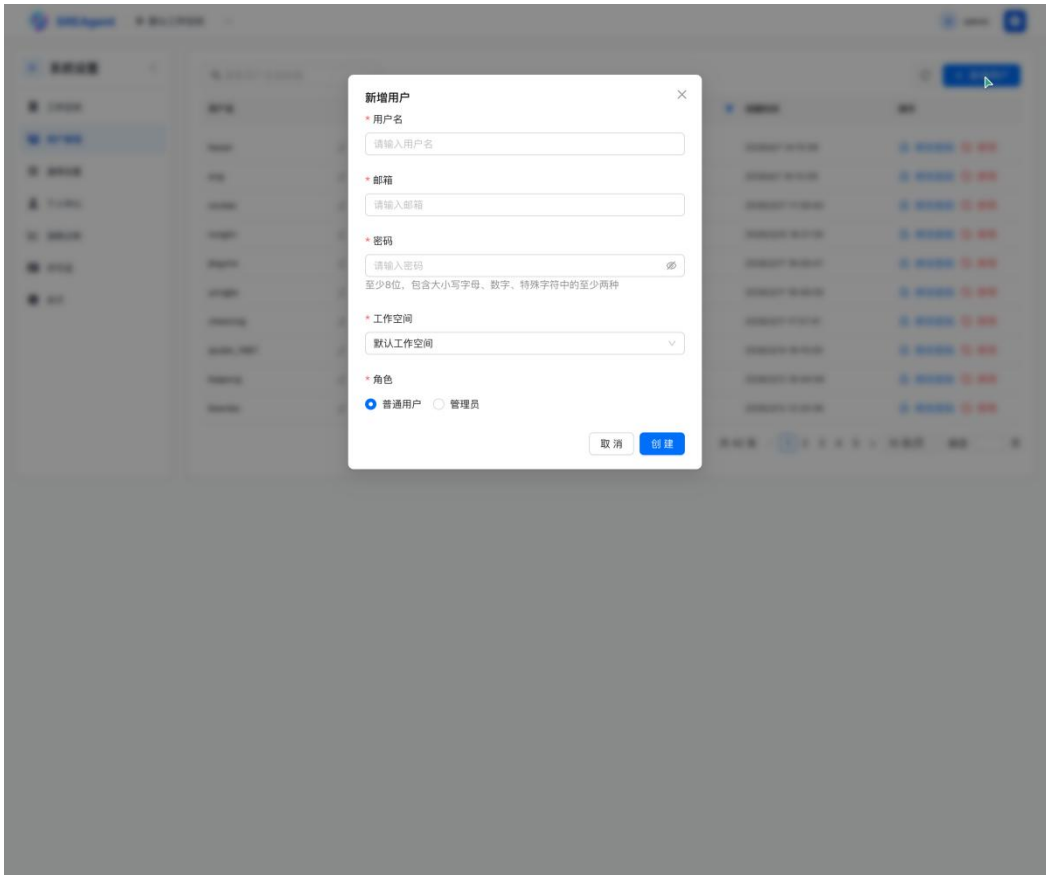
用户管理面向系统管理员，用于维护全站用户账号、角色与账号状态（启用/停用），并提供重置/修改密码等运维动作。

5.2.1 页面概览（截图）

用户列表页：



新增用户弹窗：



5.2.2 入口

- **入口**：右上角齿轮进入「系统设置」->「用户管理」。

5.2.3 权限说明

- **系统管理员**：可访问并管理用户。
- **普通用户**：通常不可见该菜单（以界面是否展示为准）。

5.2.4 页面字段说明

用户列表通常包含（以界面为准）：

- **用户名**：用于登录与展示。
- **邮箱**：用于账号识别与通知。
- **角色**：管理员 / 普通用户。
- **状态**：启用 / 停用。
- **创建时间**：审计信息。

5.2.5 核心操作指南

5.2.5.1 1) 搜索与刷新

- **搜索**：支持按“用户名或邮箱”检索用户。
- **刷新**：当你刚完成新增/改密/停用等操作后，建议刷新确认状态已生效。

5.2.5.2 2) 新增用户

点击“新增用户”，在弹窗中填写：

- **用户名**：建议简洁可读，便于审计与沟通。
- **邮箱**：用于绑定身份与后续验证（以实际策略为准）。
- **密码**：需满足平台的复杂度要求（见下文）。
- **工作空间**：可为新用户选择一个初始工作空间（以弹窗字段为准）。
- **角色**：选择管理员或普通用户。

5.2.5.3 3) 用户名与角色调整

在用户列表中通常支持行内调整：

- **修改用户名**：用于纠正命名或统一规范。
- **修改角色**：管理员/普通用户切换通常会有二次确认，避免误授权。

使用建议：

- 把管理员角色控制在最小集合，并通过成员管理/空间配置实现日常运维分权。

5.2.5.4 4) 修改密码（改密/重置）

在用户行点击“修改密码”打开弹窗，填写新密码后生效。

5.2.5.5 5) 启用/停用账号

在用户行点击“停用/启用”切换账号状态：

- **停用**：用于紧急止损或离职/离组回收权限。
- **启用**：用于恢复账号。

使用建议：

- 优先“停用”而不是“删除”，更便于审计与追溯（是否支持删除以界面为准）。

5.2.6 密码复杂度建议

平台通常会对新增/改密启用复杂度校验（以弹窗提示为准），常见要求包括：

- 长度不少于 8 位
- 大写/小写/数字/特殊字符至少包含两类
- 避免连续序列与连续重复字符

建议：在团队内统一“强密码”规范，并避免在群聊/文档中传播明文密码。

5.2.7 最佳实践

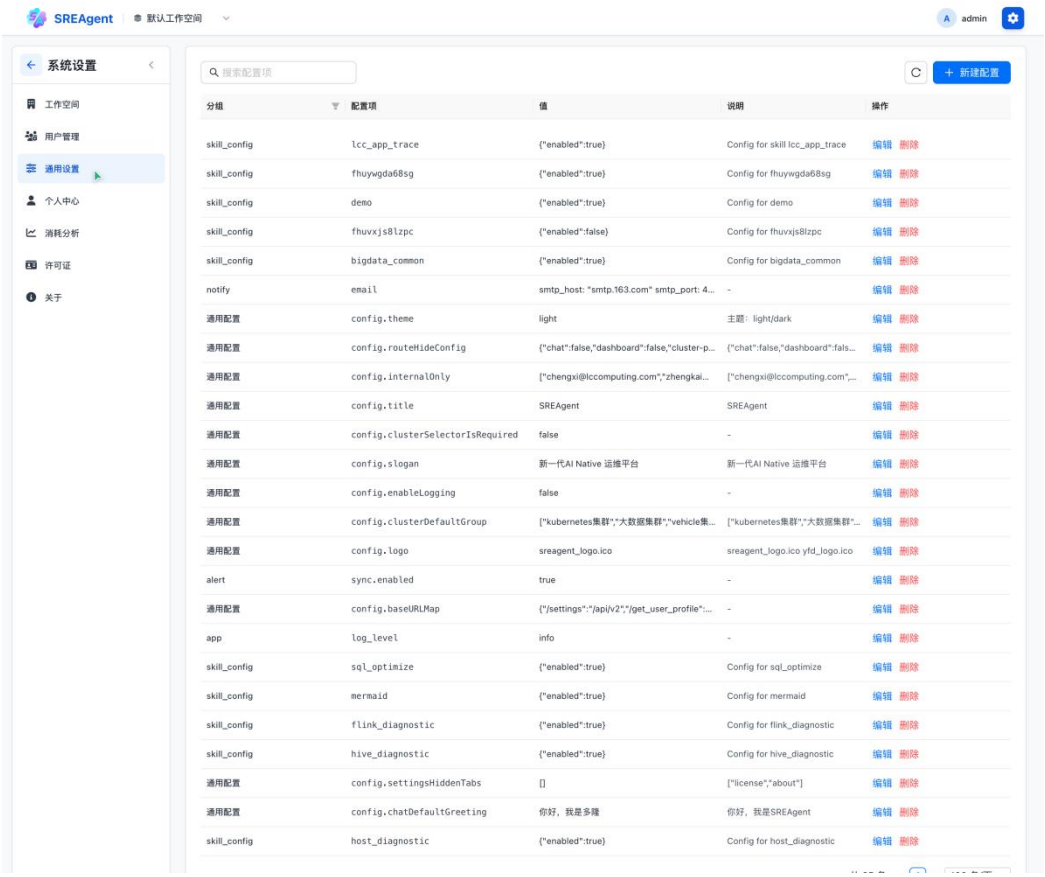
- **最小权限原则**：系统管理员仅用于全局治理；日常空间级维护建议授予空间管理员角色。
- **账号治理留痕**：对关键账号的角色变更、停用/启用建议配合工单或审批留痕。
- **保护账号策略**：如平台存在内置保护账号（不可停用/不可改名/不可降权），建议在团队内明确使用范围与交接流程。

5.3 通用设置（系统设置）

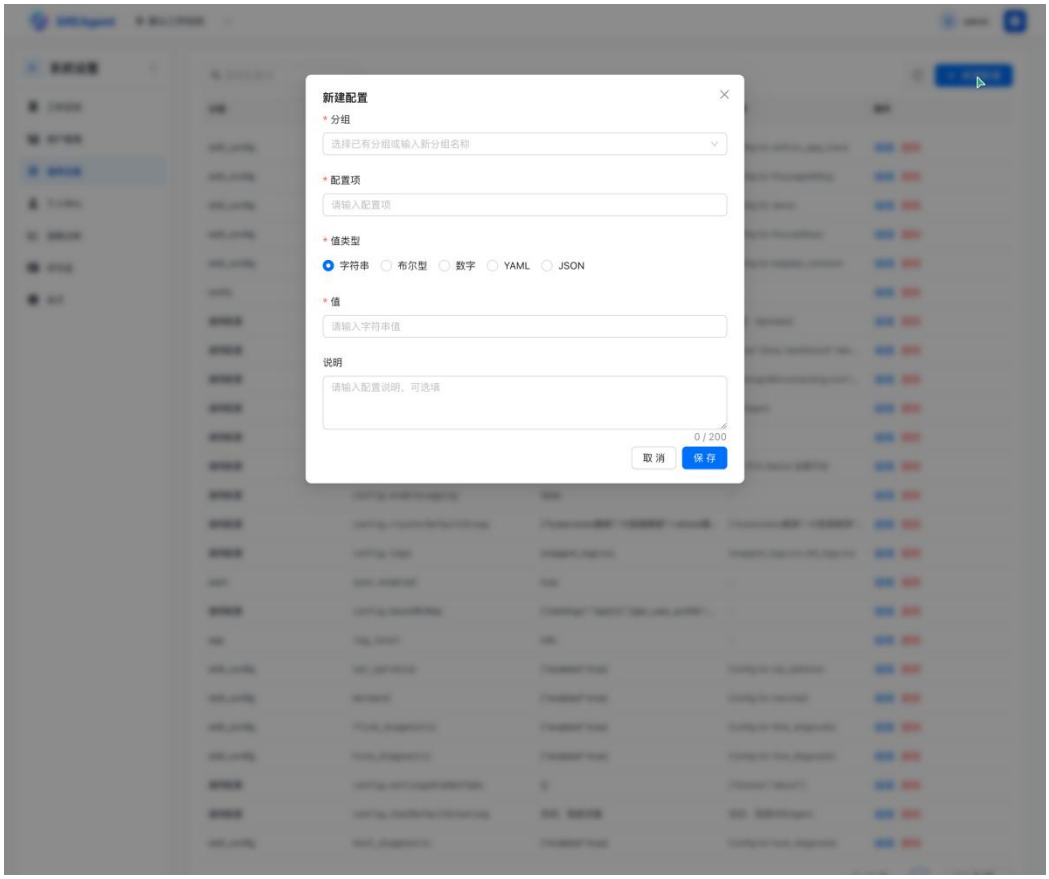
通用设置用于维护全站级别的配置项（Settings）。它更偏向“平台治理与能力开关”：同一份配置可能会影响多个模块的默认行为，因此建议仅由系统管理员维护，并配合变更流程留痕。

5.3.1 页面概览（截图）

通用设置列表页：



新建配置弹窗：



5.3.2 入口

- **入口**：右上角齿轮进入「系统设置」->「通用设置」。

5.3.3 权限说明

通用设置通常仅对系统管理员开放（以页面是否可见为准）。在免登录/无需鉴权模式下可能会放开访问，但仍建议只由少数治理人员维护。

5.3.4 数据组织方式

通用设置以“分组 + 配置项 + 值”的方式组织：

- **分组**：用于按主题归类，例如 llm、security、workspace 等。
- **配置项**：具体配置键名（建议稳定、可读、有命名规范）。
- **值类型与值**：支持 string/boolean/number/YAML/JSON；YAML/JSON 会做格式校验。
- **说明**：用于写清楚用途、影响范围、默认值口径与负责人。

5.3.5 核心操作指南

5.3.5.1 1) 搜索、筛选与分页

- **搜索**：按“配置项”关键字检索。
- **分组筛选**：快速收敛到某一类配置（例如只看安全类或模型类配置）。
- **分页**：支持切换页大小与跳页。

5.3.5.2 2) 新建配置

点击“新建配置”，按弹窗填写：

- **分组**：可选择已有分组，也可输入新分组。
- **配置项**：建议采用统一命名规范，避免含空格与特殊字符。
- **值类型**：根据配置语义选择；结构化内容优先用 YAML/JSON。
- **值**：填写后建议先自检 YAML/JSON 格式，再保存。
- **说明**：建议写清楚“影响模块 + 风险 + 回滚方式”。

5.3.5.3 3) 编辑与删除

- **编辑**：用于调整配置值与说明。建议配合变更审批/工单记录，避免误改影响线上稳定性。
- **删除**：高风险操作，通常不可恢复，且可能导致依赖模块出现默认值回退或功能异常（以实际实现为准）。

5.3.6 与“空间配置”的关系

通用设置与空间配置遵循相似的交互范式，但定位不同：

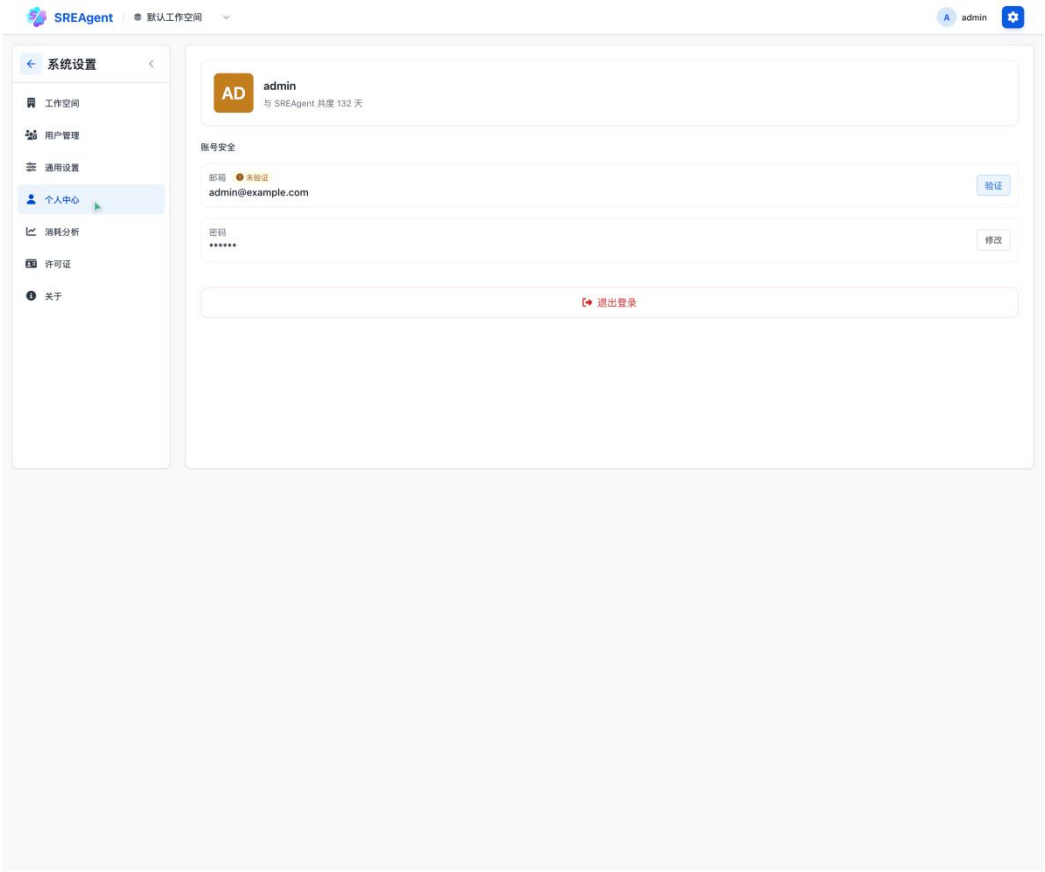
- **通用设置**：全站级，影响面更大，适合治理型配置与全局开关。
- **空间配置（集群配置/告警源配置/AGENTS.md 等）**：工作空间级，更贴近业务与团队日常使用。

5.4 个人中心（系统设置）

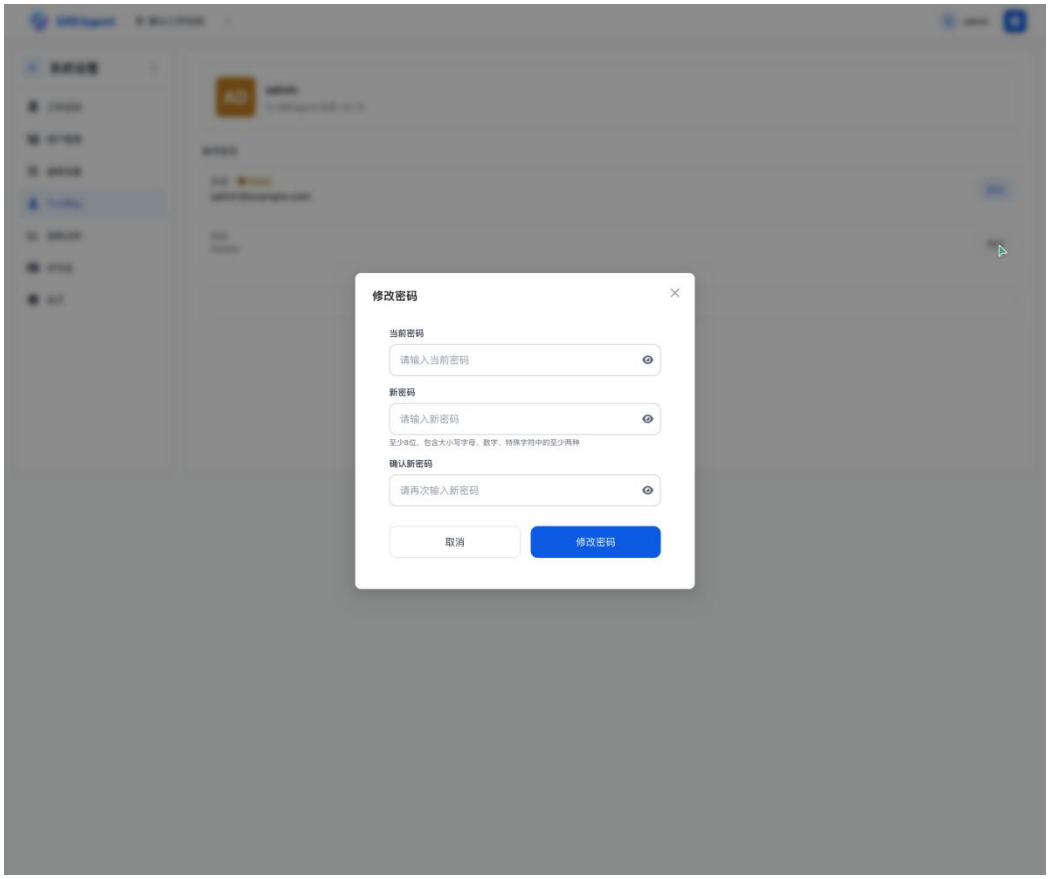
个人中心面向所有登录用户，用于查看个人信息、完成账号安全相关操作（邮箱验证、修改密码）以及退出登录。

5.4.1 页面概览（截图）

个人中心页：



修改密码弹窗：



5.4.2 入口

- **入口 1**：右上角齿轮进入「系统设置」->「个人中心」。
- **入口 2**：点击右上角用户头像菜单进入个人中心（若产品提供该入口，以界面为准）。

5.4.3 页面构成

个人中心通常包含两个区块：

- **个人信息**：头像、用户名、加入时长等。
- **账号安全**：邮箱状态、密码修改、退出登录等。

5.4.4 核心操作指南

5.4.4.1 1) 邮箱验证（如开启）

当邮箱处于“未验证”状态时，页面通常会提供“验证”按钮：

- 点击后进入邮箱验证流程（例如发送验证码/链接）。

- 验证成功后状态变更为“已验证”。

使用建议：

- 邮箱验证通常用于安全与通知能力（例如告警通知、审批通知），建议在首次登录后尽快完成。

5.4.4.2 2) 修改密码

点击“修改”打开改密弹窗，一般需要：

- 输入当前密码
- 输入新密码并再次确认

建议：

- 新密码遵循平台的复杂度提示（至少 8 位，建议包含多种字符类型）。
- 若怀疑账号泄漏，优先改密并检查近期敏感操作。

5.4.4.3 3) 退出登录

点击“退出登录”安全退出当前会话：

- 适合共享机器或临时登录场景，避免会话遗留。

5.4.5 最佳实践

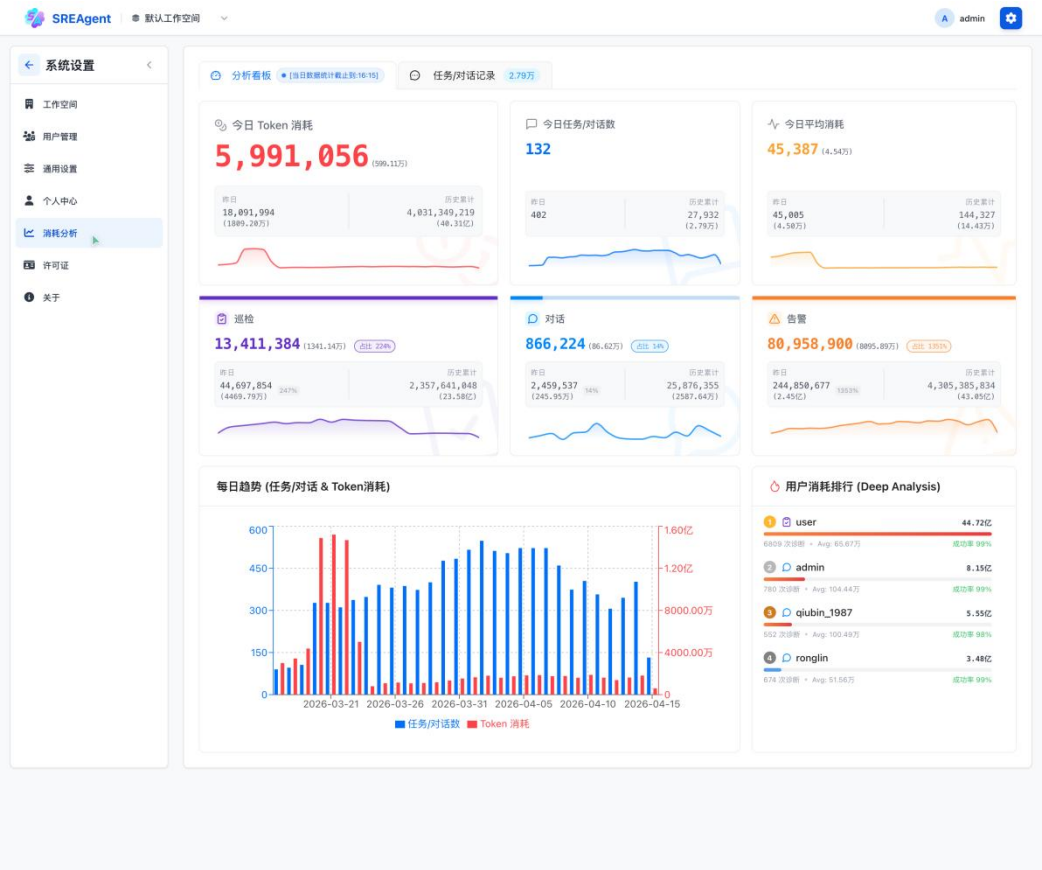
- **定期改密**：对管理员账号建议定期轮换密码，并配合双因素认证（如产品支持）。
- **安全优先**：不要在聊天工具/文档中传播明文密码，必要时使用安全渠道发放一次性密码。

5.5 消耗分析（系统设置）

消耗分析用于从“Token 消耗 / 任务与对话使用量”的视角，帮助平台管理员做成本核算、容量评估与异常用量排查。

提示：该模块通常仅对系统管理员开放，且部分部署会进一步限制为“内部账号可见”（以菜单是否展示为准）。

5.5.1 页面概览（截图）



5.5.2 入口

- **入口**：右上角齿轮进入「系统设置」->「消耗分析」。

5.5.3 页面构成

消耗分析通常包含两类视角（以界面为准）：

- **分析看板**：面向趋势与概览指标（今日消耗、任务/对话数、均值、趋势图、用户排行等）。
- **任务/对话记录**：面向明细排查与审计（按任务/会话粒度查看消耗、轮次、状态、发起人等，并支持查看详情）。

常见交互：

- 从看板点击某个指标进行“下钻”，自动带筛选条件进入明细列表。

5.5.4 核心操作指南

5.5.4.1 1) 先看趋势：判断“是日常波动还是异常突增”

建议阅读顺序：

- 先看“今日 Token / 今日任务与对话数 / 今日平均消耗”等指标，判断是否偏离常态。
- 再看趋势图（按天/时间粒度），定位异常发生的时间窗口。
- 最后看“用户消耗排行”，判断是否集中在少数用户/少数任务类型。

5.5.4.2 2) 下钻到明细：定位是谁、做了什么、消耗在哪里

在“任务/对话记录”中，常见可用信息（以界面为准）：

- **诊断目标/标题**：用于识别这条记录的业务语义。
- **发起用户**：用户名、邮箱等身份信息。
- **消耗与深度**：Token 数、消息轮次（message_count）等。
- **状态**：成功/失败/进行中/已取消等（常带过滤条件）。
- **创建时间**：用于和趋势图对齐定位。

5.5.4.3 3) 查看详情：报告 / 元数据 / 时间线（如有）

当需要进一步解释“为什么这条记录消耗高”时，详情抽屉通常会提供：

- **评估报告**：可阅读的报告内容（例如 Markdown），适合复盘与归档。
- **元数据**：JSON 形式的结构化信息，便于排查上下文与参数。
- **时间线**：按时间梳理关键事件，便于定位耗时与阶段分布。

5.5.5 最佳实践

- **建立阈值与告警**：对日均 Token、单用户消耗、单任务峰值建立阈值，异常时及时排查。
- **区分“有效消耗”和“无效消耗”**：失败率升高、重复重试、Prompt 误用通常会带来无效消耗，应优先优化。

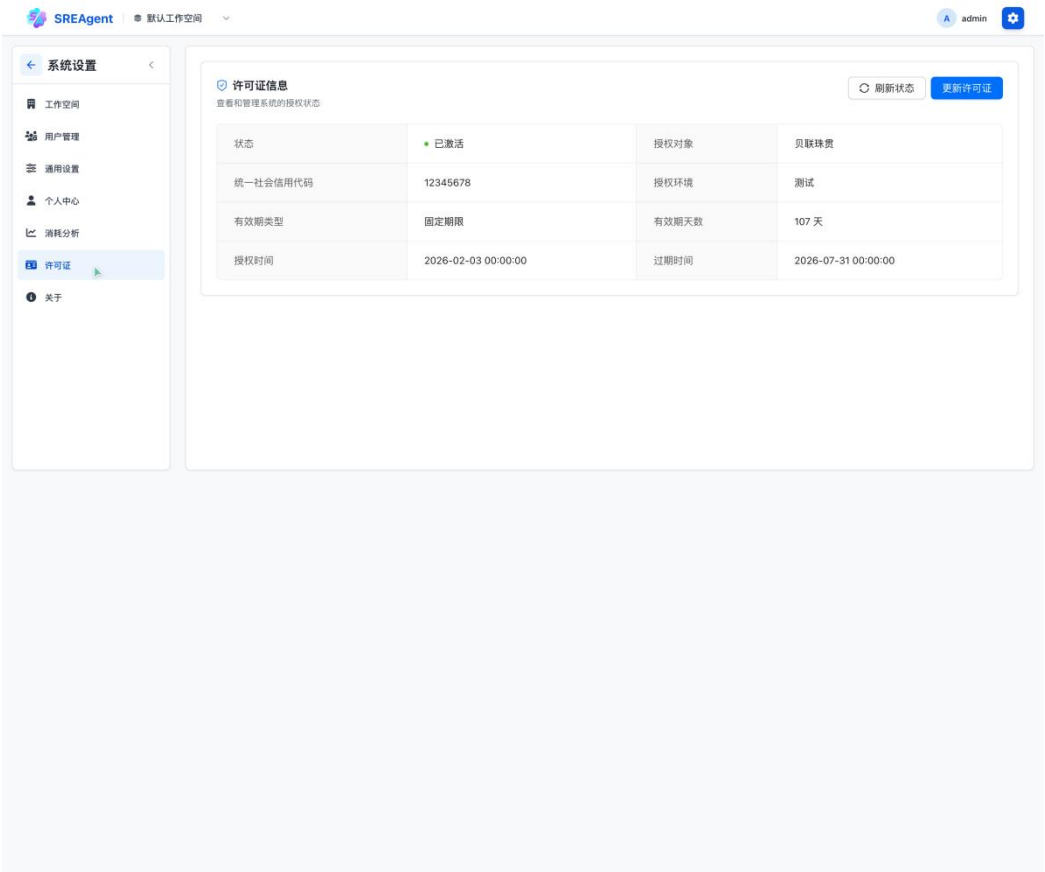
- **与治理动作联动**：对高频/高成本场景，沉淀为更高质量的 Prompt 模板、Skills 或 Runbook，降低反复试错成本。

5.6 许可证

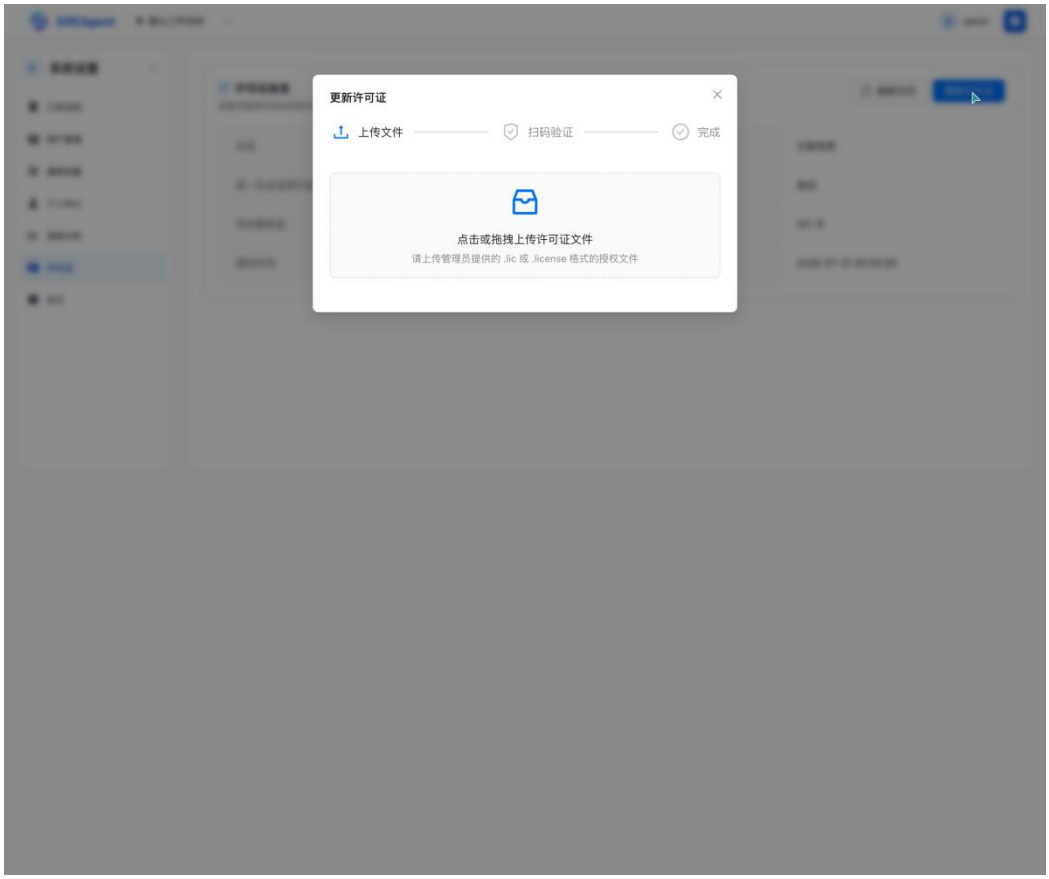
许可证用于控制系统授权状态。当许可证即将到期或已过期时，系统通常会提示风险；在到期超过一定时间后，可能进入“强制锁定”状态（以页面提示为准），并引导管理员到许可证页面处理。

5.6.1 页面概览（截图）

许可证信息页：



更新许可证（步骤 1：上传文件）：



5.6.2 入口

- **入口**：右上角齿轮进入「系统设置」->「许可证」。

5.6.3 许可证信息解读

许可证信息卡片通常会展示（以界面为准）：

- **状态**：已激活 / 即将到期 / 已过期 / 未注册等。
- **授权对象信息**：例如授权单位、授权环境等。
- **有效期信息**：授权时间、过期时间、剩余天数等。

使用建议：

- 把“过期时间”纳入运维巡检项，在临近到期前完成续签，避免影响线上使用。

5.6.4 更新/注册流程（常见为 3 步）

点击“注册许可证/更新许可证”后，通常进入三步流程（以界面为准）：

5.6.4.1 第 1 步：上传许可证文件

- 上传管理员提供的许可证文件（常见为 .lic 或 .license 格式）。
- 上传成功后系统解析文件，并进入下一步。

5.6.4.2 第 2 步：扫码验证 + 输入验证码

- 页面展示二维码。
- 管理员扫码获取验证码（常见为 6 位）。
- 在页面输入验证码并点击“验证并激活”。

5.6.4.3 第 3 步：完成

- 激活成功后展示成功提示，并刷新许可证状态。

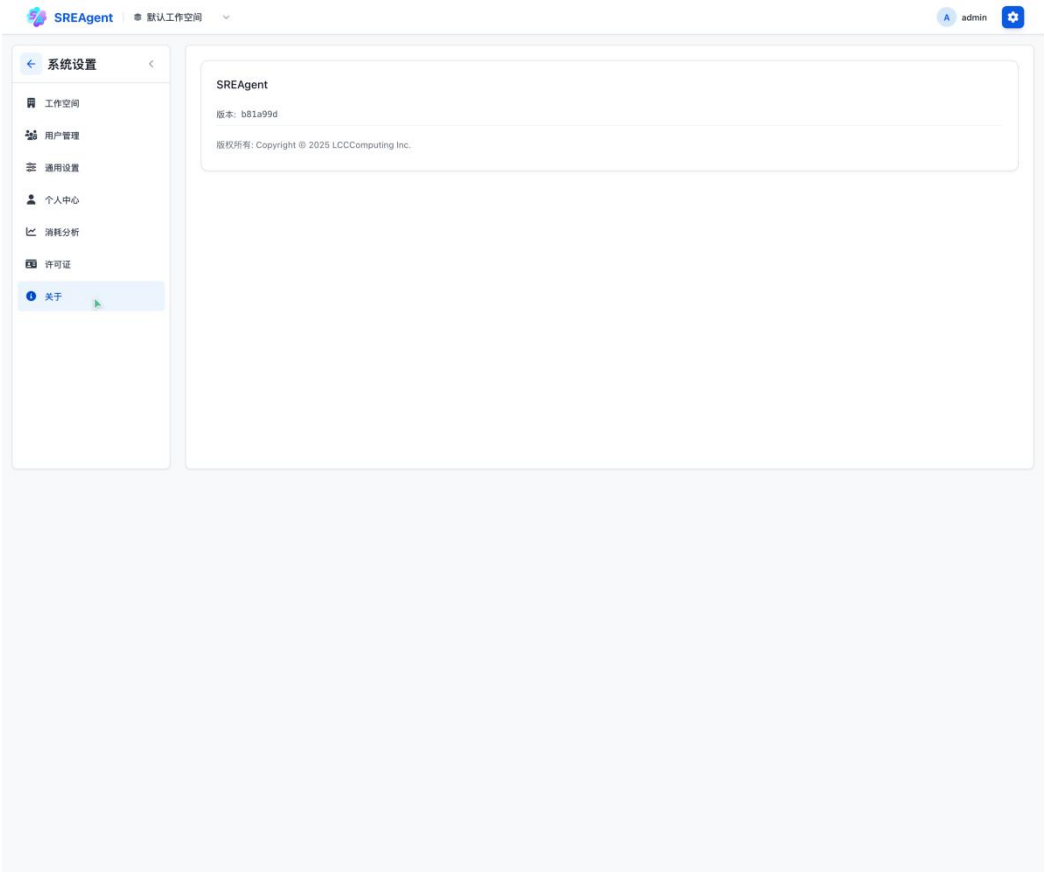
5.6.5 最佳实践

- **提前续签**：把续签当作“变更计划”提前排期，避免到期窗口期撞上业务高峰。
- **双人复核**：上传与激活建议双人复核（许可证文件来源、授权对象、有效期）。
- **留痕审计**：在团队内记录“更新时间、操作者、过期时间、许可证编号/授权对象”等关键信息。

5.7 关于

“关于”页面用于查看系统基础信息，最常用的是确认当前部署版本号与产品标识，方便定位线上问题、核对发布版本与回滚点。

5.7.1 页面概览（截图）



5.7.2 入口

- **入口**：右上角齿轮进入「系统设置」->「关于」。

5.7.3 信息说明

页面通常展示：

- **产品名称/标题**：用于确认当前访问的是哪套环境（生产/预发/测试等）。
- **版本号**：用于核对当前部署版本。
- **版权信息**：用于合规展示（以界面为准）。

5.7.4 使用建议

- 遇到线上问题或反馈缺陷时，建议在工单/群同步里带上“版本号 + 环境 + 问题截图”，减少来回确认成本。