

云容器引擎

Skill 参考

文档版本 01
发布日期 2026-06-05



版权所有 © 华为云计算技术有限公司 2026。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为云计算技术有限公司

地址：贵州省贵安新区黔中大道交兴功路华为云数据中心 邮编：550029

网址：<https://www.huaweicloud.com/>

目录

1 使用华为云云原生 Skill	1
2 华为云云原生 SKILL 最佳实践	11
2.1 使用 AI CLI 对 CCE 工作负载进行故障诊断与恢复.....	11
2.2 使用 OpenClaw 进行 CCE 集群定期巡检.....	16
2.3 使用 AI CLI 配置、查询和治理 CCE AOM 告警.....	22
2.4 基于 AI CLI 和 Skill 实现 CCE 弹性至 CCI 2.0.....	28
2.5 基于 Hermes 与飞书构建 CCE 生产环境智能运维 Agent.....	34

1 使用华为云云原生 Skill

📖 说明

本章节面向使用CCE（云容器引擎）及相关云服务的开发者、运维工程师和架构师，介绍华为云云原生Skill的能力定位、使用方式和详细参考。

Skill 概述

什么是 Skill

Skill是将专业知识、操作流程和最佳实践转化为可复用能力单元的开放能力。在AI Agent体系中，Skill用于扩展Agent的专业领域能力，使Agent能够按照预定义的流程和规则，自动执行特定领域的复杂任务。Skill的核心特点：

- 意图驱动：Agent通过读取Skill的description，自动理解何时应该触发该Skill，无需用户显式指定。
- 场景编排：单个Skill内部串联多个操作步骤，自动完成上下文收集、分析和结论输出。
- 可复用：同一个Skill可以在不同的Agent平台（Web、CLI、API）上运行，无需为每个平台单独适配。
- 可组合：多个Skill可以按 workflow 组合使用，Agent会根据任务需要自动选择和调用合适的Skill。
- 安全护栏：Skill内部定义风险约束，高危操作必须经过预览和用户确认。

华为云云原生Skill将华为云CCE、AOM、LTS、ELB、ECS、HSS等云服务的运维能力，按照故障诊断、可观测分析、巡检治理、自动恢复等场景封装为一系列Skill，使AI Agent具备专业的云原生运维能力。

适用场景

- 故障诊断：Pod CrashLoopBackOff、节点NotReady、Ingress 502、PVC Pending等。
- 可观测分析：汇聚AOM告警、LTS日志、K8s事件、Pod/Node指标形成诊断上下文。
- 巡检治理：每日集群健康检查、容量趋势预测、成本优化建议、可用性风险扫描。

- 自动恢复：扩缩容、cordon/drain节点、重启ECS、HSS漏洞修复等受控变更。
- 交付方案：容器迁移规划、资源盘点、依赖矩阵分析。
- 集群管理：CCE集群升级规划、工作负载管理、UCS集群纳管与策略治理。

安全约束与风险分级

- 核心安全约束
 - 禁止在脚本、日志、报告中输出AK/SK。
 - 所有删除、扩缩容、drain、reboot等动作必须先预览，再由用户明确确认。
 - 临时kubeconfig和证书文件必须在使用后清理。
 - 诊断类、巡检类、迁移规划类Skill只执行只读查询和报告生成。

- 风险分级机制

风险级别	示例	默认行为
R0	list/get/query/ analyze	直接执行
R1	生成报告、方案、看 板	直接执行
R2	重启异常Pod、查询 后建议	默认预览，可配置自动执行
R3	扩容、回滚、 cordon、uncordon	必须confirm=true
R4	删除集群、drain、休 眠生产集群	必须confirm=true + 强风险提示
R5	清空数据、不可逆跨 域删除	默认禁止

使用限制

- 当前Skill主要面向CCE集群及其关联云服务（AOM、LTS、ELB、ECS、HSS等）。
- 所有变更类Action默认预览模式，不自动执行。

使用说明

工作原理

Skill基于意图匹配机制工作：Agent读取Skill目录中“SKILL.md”文件头部的description，当用户输入的问题与description匹配时，Agent自动触发该Skill。Skill内部定义了完整的处理流程、可调用的工具清单和风险约束，Agent按照Skill的指引逐步执行任务。

例如，当用户问“Pod一直重启怎么办”时，Agent匹配到pod-failure-diagnoser 的description为：

```
---
name: pod-failure-diagnoser
description: Diagnose CCE Pod failures such as CrashLoopBackOff, ImagePullBackOff, OOMKilled, Pending,
```

```
Evicted, restart storms, or workload unavailable.  
---
```

Agent判断该问题与description匹配，于是自动触发pod-failure-diagnoser，执行诊断流程。

获取 Skill

华为云原生Skill通过开放仓库提供，仓库地址为[GitHub仓库](#)。

每个Skill采用自包含目录结构，包含运行该能力所需的说明和配套文件。

```
skill-name/  
├── SKILL.md      # Skill定义文件，唯一入口  
├── references/   # 参考资料  
├── scripts/      # 可执行脚本  
├── templates/   # 模板文件  
└── demo/        # 演示样例
```

安装 Skill

- 方式一：使用npm安装

```
# 安装单个Skill  
npm skills add huaweicloud/huaweicloud-skills --skill <skill-name>  
  
# 安装全部Skill  
npm skills add huaweicloud/huaweicloud-skills
```

- 方式二：从GitHub仓库手动安装

```
git clone https://github.com/huaweicloud/huaweicloud-skills.git  
  
# 安装指定Skill  
npm skills add <path>/huaweicloud-skills/skills/<skill-name>
```

不同Agent平台的加载路径和集成方式略有差异，详情请参见[平台集成示例](#)。

认证配置

使用华为云产品相关Skill前，需要根据目标云服务配置认证信息。

- 交互式配置

```
Access Key Id: <your AK>  
Secret Access Key: <your SK>
```

- 使用KooCLI配置AccessKey认证

```
hcloud configure set --cli-access-key="<your AK>" --cli-secret-key="<your SK>" --cli-mode="AKSK"
```

注意

- 仅在受信任的本地测试环境中使用明文AK/SK认证，避免凭证泄露。
- 云上环境应遵循最小权限原则，并参考[身份认证与访问控制](#)。
- 禁止将AK/SK写入脚本、日志、报告或代码仓库。

详细参考

概述

华为云云原生Skill围绕云原生资源管理和持续运维场景组织，覆盖资源生命周期、可观测与告警、故障诊断与恢复、巡检治理、解决方案交付以及多云多集群管理等能力域。

每个Skill均以独立目录提供，包含能力说明、适用场景和必要的参考资料。您可以根据业务需求选择单个Skill，也可以组合多个Skill完成跨服务、跨步骤的运维任务。以下内容按能力域列出可用Skill。

生命周期与资源管理

生命周期与资源管理包含CCE、CCI、SWR三类产品能力。产品名称仅用于分组，下表中每一行均代表一个独立的Skill。

- CCE

Skill名称	目录路径	功能描述
huawei-cloud-cce-cluster-management	skills/huawei-cloud-cce-cluster-management	管理CCE集群、节点池、节点、插件、EIP和kubecfg的全生命周期。
cce-cluster-upgrade-planner	skills/cce/cce-cluster-upgrade-planner	规划CCE Kubernetes版本升级，检查升级路径、插件兼容性、差异项和升级窗口。
cce-workload-manager	skills/cce/cce-workload-manager	管理CCE工作负载及Kubernetes资源，包括Deployment、StatefulSet、DaemonSet、Job、CronJob、HPA、Service、Ingress和配置资源。

- CCI

Skill名称	目录路径	功能描述
huawei-cloud-cci-instance-management	skills/cci/huawei-cloud-cci-instance-management	管理CCI容器实例，包括Namespace、网络、Deployment、StatefulSet、Pod、EIPPool、日志和指标。

- SWR

Skill名称	目录路径	功能描述
huawei-cloud-swr-image-management	skills/swr/huawei-cloud-swr-image-management	管理SWR命名空间、镜像仓库、标签、登录凭证和配额。

Skill名称	目录路径	功能描述
huawei-cloud-swr-image-governance	skills/swr/huawei-cloud-swr-image-governance	管理SWR权限、保留策略、共享策略、委托和不可变规则。
huawei-cloud-swr-image-automation	skills/swr/huawei-cloud-swr-image-automation	管理SWR镜像同步、触发器和自动部署流程。
huawei-cloud-swr-enterprise-instance	skills/swr/huawei-cloud-swr-enterprise-instance	管理SWR企业实例、实例内命名空间、仓库、制品、凭证、端点和域名。

可观测与智能告警

Skill名称	目录路径	功能描述
observability-context-builder	skills/observability-context-builder	汇聚AOM告警、指标、LTS日志、Pod日志和Kubernetes事件，形成诊断上下文。
alarm-correlation-engine	skills/alarm-correlation-engine	关联分析AOM active/history告警，完成去重归并、严重级别分组和告警规则核对。
log-analyzer	skills/log-analyzer	查询和分析Pod标准输出、CCE LogConfig应用日志和LTS日志。
kubernetes-event-analyzer	skills/kubernetes-event-analyzer	查询和分析Kubernetes Warning事件、重复模式及Pod、Node、Workload异常。
metric-analyzer	skills/metric-analyzer	查询和分析CCE Pod、Node及ECS、ELB、EIP、NAT指标，识别阈值异常。

故障诊断与自愈恢复

Skill名称	目录路径	功能描述
pod-failure-diagnoser	skills/pod-failure-diagnoser	诊断CrashLoopBackOff、ImagePullBackOff、OOMKilled、Pending、Evicted 和频繁重启等Pod故障。
workload-failure-diagnoser	skills/workload-failure-diagnoser	诊断Deployment、StatefulSet、DaemonSet发布失败、滚动升级卡住、副本不足和探针异常。

Skill名称	目录路径	功能描述
node-failure-diagnoser	skills/node-failure-diagnoser	诊断Node NotReady、资源压力、NPD、CNI、kubelet和容器运行时异常。
autoscaling-diagnoser	skills/autoscaling-diagnoser	诊断HPA、Cluster Autoscaler链路故障。
network-failure-diagnoser	skills/network-failure-diagnoser	诊断Service、DNS、Ingress、NetworkPolicy、ELB、EIP、NAT和VPC网络故障。
storage-failure-diagnoser	skills/storage-failure-diagnoser	诊断PVC、PV、EVS、SFS、OBS、挂载、容量和删除保护相关故障。
root-cause-analyzer	skills/root-cause-analyzer	汇总跨域证据，输出Top根因、影响范围、置信度和恢复交接。
change-impact-analyzer	skills/change-impact-analyzer	分析发布、配置、网络、安全策略和节点变更造成的故障影响。
dependency-impact-analyzer	skills/dependency-impact-analyzer	基于Service、Ingress、Pod和Node拓扑分析故障传播路径和上下游影响。
auto-remediation-runner	skills/auto-remediation-runner	生成并执行受控恢复动作，所有高风险变更默认预览并要求明确确认。

巡检、治理与持续运维

Skill名称	目录路径	功能描述
daily-cluster-inspector	skills/daily-cluster-inspector	执行周期性CCE健康检查、快速巡检和持续运维摘要。
availability-risk-scanner	skills/availability-risk-scanner	扫描高可用、AZ分布、单副本、PDB、探针、亲和性、网关和资源超配风险。
capacity-trend-forecaster	skills/capacity-trend-forecaster	分析周期性容量趋势，预测资源瓶颈，模拟HPA和节点弹性策略。
cost-optimization-advisor	skills/cost-optimization-advisor	分析空闲资源、过量Request、低利用率节点和弹性策略优化机会。
ops-report-generator	skills/ops-report-generator	汇总巡检、容量、可用性、成本和on-call上下文，生成周报、月报、SLA、容量和稳定性报告。

解决方案与交付

Skill名称	目录路径	功能描述
cce-cci-bursting-deployer	skills/cce-cci-bursting-deployer	配置、部署并验证CCE到CCI 2.0的弹性扩容能力，包括VPCEP、virtual-kubelet 和冒烟验证。
container-migration-planner	skills/container-migration-planner	盘点容器平台资源和依赖，输出迁移批次、风险和验证方案，不执行真实迁移。
全链路压测	skills/pressure-test	构建从k6客户端经ELB、nginx-ingress到业务Pod的压测链路，收集观测数据并输出性能报告。

多云、多集群管理

UCS相关Skill统一放在本分类，不再混入CCE生命周期管理。

Skill名称	目录路径	功能描述
ucs-cluster-onboarding-manager	skills/ucs/ucs-cluster-onboarding-manager	管理UCS集群纳管、生命周期、舰队分组、kubecfg和资源配额。
ucs-policy-governor	skills/ucs/ucs-policy-governor	管理UCS策略实例、策略定义、启停操作、执行状态和舰队合规审计。

使用方式

Agent根据各Skill的SKILL.md中的description自动匹配能力。需要人工定位时，先按本文档找到目标Skill，再进入对应目录查看完整说明和引用资料。

平台集成示例

在 OpenCode 中使用

OpenCode是面向终端的AI编程助手，支持通过项目目录或用户目录加载Skill。

- Skill类型

- 项目级Skill：将Skill目录放入项目根目录的skills/文件夹下。

```

my-project/
├── src/
├── skills/
│   ├── pod-failure-diagnoser/
│   │   ├── SKILL.md
│   │   ├── manifest.json
│   │   ├── skill-profile.yaml
│   │   └── references/
│   └── node-failure-diagnoser/
└── ...
    
```

OpenCode启动时会自动扫描项目目录下的skills/文件夹，加载所有Skill。您可以直接在对话中描述问题，Agent根据description自动匹配合适的Skill。

- 用户级Skill：将Skill目录放入用户配置目录。用户级Skill对所有项目生效，适合放置通用的运维Skill。

- Windows: %USERPROFILE%\opencode\skills\

- Linux/macOS: ~/.opencode/skills/

- 使用示例

```
# 进入项目目录
cd my-project

# 启动OpenCode, Skill已自动加载
opencode

# 在对话中描述问题
> 我的Pod一直在重启, 帮我看看
# Agent自动触发pod-failure-diagnoser
```

在 OpenClaw 中使用

OpenClaw是一个开源、自托管的Gateway，用于将聊天应用和渠道接入AI Agent。用户可以在本地或自有服务器上运行Gateway，并通过Skill扩展Agent能力。

OpenClaw可以从以下目录加载Skill：

目录	说明
<workspace>/skills/	当前工作区Skill，适合项目级定制。
<workspace>/agents/skills/	当前工作区内的Agent项目级Skill。
~/.agents/skills/	多个Agent可共享的Skill。
~/.openclaw/skills/	OpenClaw管理的Skill。
skills.load.extraDirs	通过配置追加的Skill目录。

OpenClaw还会加载安装时自带的Skill。将需要使用的Skill目录复制到相应的加载目录即可。例如：

```
mkdir -p ~/.agents/skills
cp -R ./skills/pod-failure-diagnoser ~/.agents/skills/
cp -R ./skills/node-failure-diagnoser ~/.agents/skills/
```

每个Skill目录应包含 SKILL.md。OpenClaw加载Skill后，Agent可以根据用户意图选择合适的Skill，并按其中定义的工作流执行任务。

关于OpenClaw的定位、Skill加载顺序和目录说明，参见[OpenClaw文档](#)和[OpenClaw Skills](#)。

在 Hermes 中使用

Hermes是面向企业级AI Agent的服务编排平台，支持通过声明式配置集成Skill。

常见问题

当描述问题时，可参考下表快速定位推荐Skill。

问题描述	推荐Skill
Pod一直重启、Pending、OOMKilled	pod-failure-diagnoser
发布失败、滚动升级卡住、副本不满足	workload-failure-diagnoser
节点NotReady、资源压力、节点漏洞	node-failure-diagnoser
HPA不扩Pod、CA不扩节点、自动弹性不生效	autoscaling-diagnoser
Ingress 502、Service不通、ELB链路异常	network-failure-diagnoser
PVC Pending、FailedMount、容量耗尽	storage-failure-diagnoser
CCE告警很多，需要合并分析	alarm-correlation-engine
查询Pod标准输出或LTS应用日志	log-analyzer
分析Kubernetes事件趋势	kubernetes-event-analyzer
查询CCE Pod/Node指标和资源使用排名	metric-analyzer
需要汇聚日志、事件、指标、告警	observability-context-builder
业务不可用，需要综合根因分析	root-cause-analyzer
发布、配置、网络、安全策略或节点变更后出现故障	change-impact-analyzer
某个服务故障会影响哪些入口和上下游	dependency-impact-analyzer
需要扩容、重启、drain、漏洞修复	auto-remediation-runner
做每日巡检或周期性健康检查	daily-cluster-inspector
做成本优化、Request过量分析	cost-optimization-advisor
做容量趋势预测、弹性模拟	capacity-trend-forecaster
做可用性风险扫描、PDB/探针检查	availability-risk-scanner

问题描述	推荐Skill
做周报、月报、SLA运维报告	ops-report-generator
做容器迁移方案和资源盘点	container-migration-planner
CCE到CCI弹性扩容配置	cce-cci-bursting-deployer
CCE集群版本升级规划	cce-cluster-upgrade-planner
CCE/UCS工作负载管理	cce-workload-manager
UCS集群纳管和舰队管理	ucs-cluster-onboarding-manager
UCS策略治理和合规审计	ucs-policy-governor
SWR镜像生命周期管理	huawei-cloud-swr-image-management
SWR镜像治理	huawei-cloud-swr-image-governance
SWR镜像自动化	huawei-cloud-swr-image-automation
压力测试方案和执行	全链路压测

相关文档

文档	说明	路径
CCE产品资料	云容器引擎CCE官方文档	华为云CCE文档
Skill开放仓库	华为云云原生Skill代码仓库	huaweicloud/huaweicloud-skills

2 华为云云原生 SKILL 最佳实践

2.1 使用 AI CLI 对 CCE 工作负载进行故障诊断与恢复

应用场景

在CCE集群中，工作负载发布、扩缩容或配置变更后，可能出现Pod长时间未就绪、Deployment滚动升级卡住、容器频繁重启、镜像拉取失败、调度失败、服务端点异常等问题。传统排障通常需要运维人员在Deployment、ReplicaSet、Pod、Events、容器日志、探针配置和监控数据之间反复切换，定位链路长、人工判断成本高。

通过AI CLI结合云原生Skill能力，用户可以使用自然语言描述故障现象，由Agent自动完成上下文识别、证据采集、根因分析、恢复方案预览、用户确认、恢复执行和结果验证，帮助运维团队把工作负载故障处理沉淀为可重复的标准流程。

本实践面向CCE工作负载故障诊断与恢复的通用场景，不绑定固定命名空间、工作负载名称或单次测试环境。文中的 ai-diagnose-demo 仅表示演示命名空间，实际使用时请替换为您的目标集群和业务命名空间。

约束与限制

- AI CLI的诊断准确性依赖于集群访问权限、日志保留时间、事件完整性和可观测数据质量。
- 所有写操作必须遵循“预览 + 确认”模式，禁止在未确认的情况下执行回滚、重启、扩缩容或配置修改。
- 生产环境中，建议将诊断权限和恢复权限分离，并通过审计日志记录AI CLI的所有操作。
- 回滚操作依赖于工作负载的历史Revision。如果历史版本已清理，需要采用重新发布或配置修复等方式恢复。
- 如果故障涉及数据库变更、数据格式变更、外部依赖或消息堆积，回滚前需要评估数据一致性和业务补偿方案。
- 对于多集群、多命名空间或多异常对象的场景，AI CLI应要求用户确认目标范围，避免跨业务的误操作。
- 禁止在提示词、诊断报告或恢复预览中输出AK/SK、Token、证书、真实项目ID等敏感信息。

- 建议先在演示或测试命名空间验证Skill流程，确保流程的正确性和安全性，再推广到生产集群。

前提条件

- 已创建CCE集群，目标工作负载已部署。
- 已安装或接入AI CLI，并完成相关Skill注册。
- AI CLI具备读取工作负载、Pod、Events、日志、版本历史和服务端点的权限。
- 如需执行恢复操作，AI CLI还需具备回滚、重启、扩缩容或修改工作负载配置的权限。

涉及的 Skill

Skill名称	功能描述
huawei-cloud-cce-workload-failure-diagnoser	采集工作负载、Pod、Events、日志和版本历史，输出诊断结论。
huawei-cloud-cce-auto-remediation-runner	生成恢复预览，并在用户确认后执行回滚、重启、扩缩容等动作。

操作流程

步骤一：启动 AI CLI 会话

根据企业实际接入方式启动AI CLI。若使用命令行方式，可进入交互会话：

```
aicli chat
```

如果AI CLI已接入运维平台、ChatOps工具或流水线，也可以直接在对应入口发起自然语言请求。

步骤二：使用自然语言描述故障

用户不需要手动拼接多条Kubernetes命令，只需描述目标集群、命名空间和故障现象。建议在输入中说明是否允许执行恢复操作，以及恢复前是否需要预览。例如，提示词如下。

请帮我诊断北京四cce-ai-ops-demo集群中ai-diagnose-demo命名空间的工作负载发布异常，现象是最近一次更新后Pod长时间未就绪。请先分析根因并给出恢复建议，恢复操作执行前请让我确认。

如果不确定异常对象，可以让AI CLI先扫描命名空间。

请先检查cce-ai-ops-demo集群中ai-diagnose-demo命名空间的异常工作负载，找出发布失败或Pod未就绪的对象，并输出诊断结论和恢复建议。

步骤三：确认诊断范围

AI CLI会根据用户输入识别区域、集群、命名空间、资源类型、故障时间窗口和故障现象。若命名空间中存在多个异常工作负载，AI CLI应先列出候选对象及异常摘要，让用户确认诊断范围。

建议确认以下信息。

信息	说明
目标集群	支持通过集群名称、集群ID或区域+集群名称识别。
命名空间	用于限定诊断范围，避免跨业务误操作。
工作负载类型	可诊断Deployment、StatefulSet、DaemonSet等常见工作负载。
故障现象	例如Pod未就绪、滚动升级卡住、容器重启、镜像拉取失败、调度失败。
恢复边界	是否允许执行回滚、重启、扩缩容或配置修改等动作。

步骤四：自动采集诊断证据

AI CLI调用工作负载故障诊断Skill后，应按照从控制面到数据面的顺序采集证据，避免只依赖单点日志或单条事件作出判断。

诊断维度	关键证据	诊断价值
工作负载状态	期望副本、就绪副本、可用副本、更新副本、状态条件	判断发布是否完成、可用性是否受影响。
版本关系	Revision历史、ReplicaSet状态、新旧版本副本分布	判断是否存在健康历史版本，是否具备回滚条件。
Pod生命周期	Pending、Running、Ready、RestartCount、容器状态	判断故障发生在调度、启动、运行还是就绪阶段。
Kubernetes事件	FailedScheduling、FailedPull、Unhealthy、BackOff等事件	快速定位调度、镜像、探针、容器启动类问题。
容器日志	最近异常日志、启动日志、健康检查相关日志	识别应用内部错误、依赖异常或配置问题。
探针配置	readinessProbe、livenessProbe、startupProbe配置和结果	判断健康检查路径、端口、协议和超时配置是否合理。
服务端点	Service、EndpointSlice、Ingress或负载均衡后端	判断故障是否影响业务流量接入。
可观测数据	指标、告警、日志趋势	识别资源压力、异常流量、依赖抖动等外部因素。

步骤五：输出诊断结论和恢复建议

诊断完成后，AI CLI应输出结构化结论，帮助用户快速判断是否需要恢复操作。

建议输出以下内容。

输出项	内容
诊断结论	明确故障类型，例如探针失败、镜像拉取失败、容器启动失败、调度失败、资源不足或配置异常。
影响范围	说明受影响的命名空间、工作负载类型、不可用副本数量和业务访问影响。
关键证据	列出支撑结论的状态条件、事件、日志或指标，不输出敏感信息。
根因分析	说明故障发生在发布链路的哪个阶段，以及为什么会导导致工作负载不可用。
恢复建议	给出一个或多个恢复选项，并说明适用场景、风险等级和预期效果。
是否需要确认	对回滚、重启、扩缩容、配置修改等变更动作标记为需要用户确认。

步骤六：预览恢复方案

如果用户希望AI CLI继续处理故障，AI CLI应调用自动恢复Skill生成恢复预览。预览阶段只展示计划，不改变集群状态。

常见恢复动作如下所示。

恢复动作	适用场景	风险控制
回滚到健康 Revision	新版本发布后不可用，旧版本仍可稳定承载业务。	校验历史版本、当前可用副本和回滚影响范围。
重启异常 Pod	单个Pod进入异常状态，工作负载模板本身未发现明显配置问题。	避免批量重启导致可用副本不足。
临时扩容	当前可用副本不足，业务需要先恢复容量。	评估资源配额、节点容量和HPA策略。
修正工作负载配置	探针、环境变量、镜像Tag、Secret或ConfigMap引用存在错误。	展示配置差异，确认变更窗口和回退方式。
暂停或恢复发布	滚动升级异常，需要阻止继续扩大影响或继续完成发布。	明确暂停/恢复后的发布状态和人工后续动作。

恢复预览建议包含以下信息：

- 将执行的动作和目标资源范围。
- 变更前后的关键配置或状态差异。
- 风险等级、业务影响和是否需要变更窗口。
- 验证方式和回退路径。
- 用户需要确认的具体语句。

步骤七：确认并执行恢复

用户确认恢复计划后，AI CLI才可以调用自动恢复Skill执行动作。建议用户使用明确表达，例如：

确认按预览方案执行恢复。

执行过程中，AI CLI应持续反馈变更状态。如果恢复动作失败，应停止后续变更，输出失败原因、已执行动作、当前集群状态和建议的人工处理方式。

步骤八：验证恢复结果

恢复完成后，AI CLI需要再次读取工作负载状态和相关证据，确认故障是否真正恢复。

建议验证以下内容。

验证项	期望结果
工作负载状态	期望副本、就绪副本、可用副本符合预期，发布状态稳定。
Pod状态	Pod处于Running且Ready，无持续重启、拉取失败或调度失败。
事件和日志	不再出现同类高频异常事件，容器日志无新的关键错误。
服务端点	Service或EndpointSlice具备可用后端，业务流量可被正确转发。
业务探测	如已配置健康检查或外部探测，探测结果恢复正常。
告警状态	相关告警恢复或进入收敛状态，无新的高风险告警触发。

Skill 执行流程

阶段	输入	Skill动作	输出
目标识别	自然语言中的区域、集群、命名空间、故障现象	解析诊断范围，补全缺失信息，必要时请求用户确认	明确的诊断目标和边界
证据采集	目标工作负载和时间窗口	查询工作负载、Pod、Events、日志、版本历史和服务端点	多维度诊断证据
根因分析	诊断证据	识别故障阶段，排除不匹配原因，输出置信度	根因结论和关键证据
恢复规划	根因、影响范围、权限边界	生成回滚、重启、扩缩容或配置修复预览	恢复计划、风险等级、验证方式
用户确认	用户确认语句	检查确认内容是否匹配预览计划	可执行恢复任务
恢复执行	已确认的恢复任务	调用Kubernetes API或CCE相关接口执行变更	执行结果和中间状态
恢复验证	恢复后的工作负载状态	复查状态、事件、日志、端点和业务探测	最终结论和后续建议

预期结果

完成本实践后，用户可以通过AI CLI自然语言对话完成以下闭环：

1. 识别目标CCE集群、命名空间和异常工作负载。
2. 自动汇聚工作负载状态、Pod状态、版本历史、事件、日志、探针和服务端点等证据。
3. 输出可解释的根因分析，而不是只给出单条命令结果。
4. 对恢复动作进行预览，明确风险、影响范围和验证方式。
5. 在用户确认后执行受控恢复动作。
6. 自动验证恢复结果，并输出后续修复建议。

后续建议

为了提升CCE工作负载故障处理效率，建议结合本实践沉淀以下能力：

- 标准化提示词
在团队SOP中固化诊断提示词模板，明确集群、命名空间、故障现象、时间窗口和恢复边界。
- 完善发布前检查
在CI/CD流水线中增加镜像启动检查、探针路径校验、配置引用校验和基础冒烟测试。
- 配置合理发布策略
对关键业务配置合适的滚动更新策略、PDB、HPA和Revision保留数量，降低发布异常对业务的影响。
- 增强可观测能力
接入日志、指标、事件和告警数据，让AI CLI能够从更多维度判断根因。
- 建立变更审计
对AI CLI触发的恢复动作记录操作者、确认语句、执行动作、执行结果和恢复验证结论，便于复盘和合规审计。

相关文档

- [CCE最佳实践汇总](#)
- [容器应用管理](#)
- [Kubernetes Deployments](#)
- [Kubernetes Pod Lifecycle](#)

2.2 使用 OpenClaw 进行 CCE 集群定期巡检

应用场景

生产环境中的CCE集群需要持续关注节点健康、Pod状态、核心插件、资源利用率、Kubernetes事件、AOM告警和业务入口状态。通过OpenClaw Agent对接集群巡检能力，可以用自然语言配置周期性巡检任务，让Agent自动完成集群健康检查、告警聚合、异常分析、风险分级、报告生成和通知推送。

本实践推荐采用“先快检、异常再深检”的方式运行巡检：

- 集群状态正常时，Agent输出简洁的健康摘要，减少无效噪音。
- 快检发现异常时，Agent自动扩展到Pod、Node、Event、AOM、ELB和资源利用率等维度。
- Agent会查询近24小时AOM告警，并按告警类型、严重程度、当前状态和重复频次进行聚合，区分活跃告警、已恢复告警、突发告警和常态重复告警。
- 深度巡检会补充Pod历史指标TopN，以及节点CPU、内存、磁盘TopN，帮助判断异常是否与资源水位相关。
- AI会基于巡检证据进行风险分级，并在报告中呈现影响范围、可能原因和下一步建议；分级结果用于摘要和建议，不要求工具返回固定字段。
- 巡检过程只执行只读查询和报告生成，不会自动执行扩容、删除、重启、drain等变更动作。

通过OpenClaw Agent可以实现：

- 每天或每周定时执行CCE集群巡检。
- 自动生成Markdown和HTML巡检报告。
- 将巡检摘要和报告链接通过邮件推送给运维团队。
- 归档历史巡检报告，便于趋势对比和复盘。
- 在发现严重风险时，继续转交相关诊断能力做深入分析。

约束与限制

- 巡检过程只执行只读查询和报告生成，不会自动执行修复动作。
- 巡检任务执行期间会调用CCE、AOM、LTS、ELB等云服务的API，会产生少量API调用费用或日志查询成本。
- 报告存储在OBS中，会产生相应的存储费用。
- 邮件发送频率受SMTP服务或华为云邮件推送服务配额限制，建议合理设置巡检频率。
- 严禁将AK/SK、Token、证书及真实项目ID写入文档、代码或对话输出。

注意事项

- 深度巡检会采集更多指标和上下文，包括近24小时AOM告警、Pod历史指标TopN、节点CPU/内存/磁盘TopN，执行时间可能明显长于快检。
- Pod资源TopN基于历史指标时间窗查询，结果可能包含查询时间窗内出现过但当前已不存在的Pod；可结合当前Pod列表判断对象是否仍存在。
- 风险分级由AI根据工具返回的事实证据生成，建议先查看关联事件、日志和指标，再决定是否进入恢复流程。

前提条件

- 已创建CCE集群且状态为“运行中”。
- 已开通OpenClaw服务并完成Agent初始化。
- Agent已完成华为云云原生能力接入。
- 目标CCE集群已安装云原生监控插件。
- 目标CCE集群已按照最佳实践配置AOM告警规则，详情请参见[使用AI CLI配置、查询和治理CCE AOM告警](#)。

- 已配置华为云访问凭证，建议通过OpenClaw密钥管理或环境变量注入，不在文档、脚本或对话中暴露AK/SK。
- 巡检账号已具备CCE、AOM、LTS、ELB等相关资源的只读查询权限。
- 如需邮件通知，已准备SMTP服务或华为云邮件推送服务。
- 如需报告归档，已准备OBS桶或其他报告存储位置。

推荐输入

可以直接在OpenClaw对话中描述目标集群、巡检周期、巡检范围和通知方式。

使用场景及推荐输入

使用场景	推荐输入
创建每日巡检	帮我为北京四 test-ai-diagnoses 集群创建每日巡检任务，每天上午9点执行，先做快检，发现异常再做深度巡检，并把报告发送给运维组。
立即执行巡检	立即巡检北京四 test-ai-diagnoses 集群，先做快检，发现异常再做深度诊断。
查看最近报告	查看 test-ai-diagnoses 集群最近一次巡检报告，并按严重程度列出风险。
分析一周趋势	汇总 test-ai-diagnoses 集群最近7天巡检结果，告诉我风险是否变多。
深入分析异常	继续分析巡检报告中的高风险节点问题，关联事件、指标和相关Pod。

巡检结果建议重点查看以下内容

输出内容	关注点
巡检结论	集群整体是否健康，是否存在高风险异常。
异常分组	异常集中在Pod、Node、Event、AOM、ELB还是Resource。
影响范围	影响哪些命名空间、节点、工作负载或业务入口。
风险趋势	与前一天或最近7天相比，问题是否新增、扩大或恢复。
建议动作	继续观察、进入专项诊断、扩容评估、规则优化或转交恢复流程。

操作流程

步骤一：创建 CCE 集群周期巡检任务

让Agent按指定周期创建CCE集群巡检任务。Agent会根据输入识别区域、集群名称、巡检时间、报告格式和通知方式，并生成巡检计划。

1. 在OpenClaw对话中输入以下内容。
帮我为北京四 test-ai-diagnoses 集群创建每日巡检任务，每天上午9点执行，先做快检，发现异常再做深度巡检，生成Markdown/HTML巡检报告，并发送给ops-team@company.com。
2. Agent会自动生成巡检计划，确认以下内容。

配置项	示例	说明
区域	cn-north-4	目标CCE集群所在区域
集群名称	test-ai-diagnoses	需要巡检的CCE集群
执行周期	每天09:00	建议选择业务低峰或值班交接前
巡检策略	先快检，异常后深检	默认减少不必要的重型检查
报告格式	Markdown、HTML	便于邮件阅读和历史归档
通知对象	ops-team@company.com	接收巡检摘要和报告链接
存储位置	obs://your-bucket/reports/	保存历史巡检报告

3. 任务生成后，Agent会按计划执行巡检。建议立即触发一次巡检，用于验证配置是否正确。
立即执行一次test-ai-diagnoses集群巡检，并发送测试报告。

步骤二：查看巡检报告并识别风险

1. 巡检完成后，Agent会生成巡检摘要和完整报告。可以直接查看最近一次巡检结果。
查看 test-ai-diagnoses 集群最近一次巡检报告，并按严重程度列出风险。
2. Agent会先返回巡检摘要。

巡检项	示例结果	关注点
巡检结论	Warning	集群是否存在需要处理的风险
检查项数量	12项，通过9项，警告2项，失败1项	是否有新增异常
节点健康	3/3节点正常	是否存在NotReady、资源压力或节点事件

巡检项	示例结果	关注点
Pod状态	2个Pod异常	是否存在CrashLoopBackOff、Pending、Evicted
AOM告警	近24小时140条，4条未恢复	是否存在持续告警、突发告警或常态重复告警
核心插件	正常	CoreDNS、网络、存储等插件是否健康
Pod资源TopN	近24小时CPU/内存TopN	是否存在历史高水位Pod或已消失Pod的指标痕迹
节点资源TopN	节点CPU/内存/磁盘TopN	是否存在节点CPU、内存、磁盘容量风险

3. 当巡检发现异常时，Agent会基于工具返回的事实证据进行严重程度判断，并输出问题清单。

严重程度	分类	资源	问题	证据	建议
高	Pod健康	default/test	可用副本为0	Deployment期望2个副本，当前Ready为0；相关Pod处于Pending	查看Pod事件和镜像拉取状态，优先恢复工作负载可用性
高	Node资源	192.168.32.2	节点CPU持续高位	节点CPU当前100%，近24小时窗口内持续高位	定位该节点上的高CPU进程或Pod，必要时评估迁移或扩容
中	AOM告警	default/test-*	镜像拉取失败重复告警	近24小时出现FailedPullImage、BackOffPullImage，仍有未恢复告警	修正镜像地址或版本，修复后重新触发Deployment滚动更新

4. 对于严重或反复出现的问题，可以让Agent继续分析。

继续分析 default/test 副本不可用问题，关联近24小时AOM告警、Kubernetes事件、Pod状态和相关指标。

Agent会在巡检报告基础上继续汇聚上下文，并输出根因线索、影响范围和下一步处理建议。例如，针对严重告警或高风险资源异常，可以让Agent进一步分析对应时间窗内的相关Pod、节点指标、事件和日志，判断异常是否具备同一根因。

步骤三：查看历史趋势和报告归档

周期巡检的价值不只是发现当天异常，也包括观察风险是否持续、扩大或恢复。可以让Agent汇总一段时间内的巡检结果：

汇总 test-ai-diagnoses 集群最近7天巡检结果，按日期列出高、中、低风险变化和新增问题。

Agent可以输出趋势摘要。

日期	执行状态	总检查项	高风险	中风险	低风险	新增问题	备注
2026-05-31	成功	12	1	2	1	1	Pod重启问题新增
2026-05-30	成功	12	1	1	1	0	节点内存压力持续
2026-05-29	成功	12	2	2	2	2	出现核心插件异常
2026-05-28	成功	12	0	0	0	0	集群健康

也可以查看报告归档位置。

列出 test-ai-diagnoses 集群最近7天巡检报告的Markdown和HTML链接。

报告预期将保留以下内容。

报告内容	说明
巡检摘要	集群整体状态、检查项数量、高/中/低风险数量
异常清单	按Pod、Node、Event、AOM、ELB、Resource分类展示
风险趋势	与前一次巡检或近7天趋势对比
根因线索	对严重异常给出相关日志、事件、指标入口
建议动作	继续观察、进入专项诊断、容量评估或转交恢复流程

预期结果

完成本实践后，OpenClaw Agent能够帮助完成以下闭环：

1. 按指定周期自动执行CCE集群巡检。
2. 默认先执行快检，发现异常后再进入深度诊断或并行巡检。
3. 按Pod、Node、Event、AOM、ELB、Resource等维度汇总异常。
4. 由AI基于巡检证据标记风险严重程度和影响范围。
5. 自动生成Markdown和HTML巡检报告，并通过邮件推送给运维团队。
6. 将历史巡检报告归档，支持按天查看和趋势对比。
7. 对严重或持续风险，继续联动日志、事件、指标和相关诊断能力做根因分析。

常见问题

每次巡检都需要深度巡检吗？

不建议。推荐默认先执行快检，只有发现异常时再进入深度诊断或并行巡检。这样可以减少不必要的API调用、日志查询和报告噪音，提高巡检效率。

巡检发现高风险问题后会自动修复吗？

不会。OpenClaw Agent在本实践中只做巡检和报告生成，不执行修复动作。确认需要恢复时，可以让Agent转交对应诊断或恢复能力，并在执行任何变更前进行确认。

为什么 Pod 资源 TopN 中会出现当前不可见的 Pod？

Pod资源TopN用于分析历史时间窗内的资源使用情况。默认查询近24小时的历史指标，因此可能出现已经删除或重建过的Pod。可以让Agent继续查询当前Pod列表，并将历史指标对象与当前存量对象分开说明，以便更好地理解资源使用情况。

为什么收不到邮件？

建议检查邮件收件人、SMTP或邮件推送服务配置、发送记录、邮件服务配额和企业邮箱拦截策略。如果报告已生成但邮件未发送成功，可以先通过OpenClaw控制台或报告归档路径查看报告。

历史报告应该保留多久？

建议生产集群至少保留30天巡检报告。如果需要月度稳定性复盘、SLA统计或容量趋势分析，可以保留90天或更长时间，并配置OBS生命周期策略控制存储成本。这样可以确保在需要时能够快速访问历史数据，同时有效管理存储成本。

相关文档

- [云容器引擎CCE](#)：查询CCE集群、节点、工作负载、插件和云原生观测相关产品说明。
- [应用运维管理AOM](#)：查询AOM指标、告警、日志和应用观测相关说明。

2.3 使用 AI CLI 配置、查询和治理 CCE AOM 告警

应用场景

CCE集群上线后，运维团队通常需要尽快完成告警规则初始化，并在日常排障中持续关注当前活跃告警、历史告警、恢复记录和通知规则。通过AI CLI Agent对接alarm-correlation-engine Skill，您可以用自然语言完成CCE AOM告警规则配置、告警查询、告警聚合分析和严重告警根因追查。

相比只查看单条告警列表，本实践推荐客户按时间窗分析集群告警。AI CLI会自动归并同类告警，标记严重程度，区分常态告警、突发告警和持续未恢复告警，帮助客户快速判断“哪些告警最紧急、影响哪些资源、下一步应该查什么”。

通过AI CLI可以实现：

- 一键为指定CCE集群创建AOM告警中心推荐规则。

- 未指定通知规则时，自动创建集群级默认通知规则。
- 创建前预览规则数量、规则类型、通知方式和缺失参数，客户确认后再执行。
- 创建完成后自动查询告警规则列表，确认50个规则已成功创建。
- 按集群和时间窗查询当前活跃告警、历史告警和恢复记录。
- 对告警进行去重、分组、严重程度标记、突发/持续/常态识别和根因线索分析。

本实践以北京四test-ai-diagnoses集群批量配置CCE AOM告警规则，并分析最近4小时告警态势为例，演示如何通过自然语言完成告警配置、告警查询和聚合分析。

约束与限制

- 可不指定已有通知规则，未指定时AI CLI将自动调用集群级默认通知规则。
- 自动创建通知规则时，必须提供可用的SMN主题名称或主题URN。
- 指标告警依赖目标集群已绑定AOM CCE Prometheus实例。
- 事件告警依赖CCE事件上报链路，建议同步确认日志采集与节点故障检测插件状态正常。
- 创建告警规则须经客户明确确认后方可执行。
- 严重程度标记仅作排障辅助，不可替代生产变更审批与人工确认机制。
- 分析告警时建议显式指定时间窗，如最近30分钟、最近4小时或特定起止时间段。
- 严禁将AK/SK、Token、证书及真实项目ID写入文档、代码或对话输出。

前提条件

- 已在AI CLI中注册alarm-correlation-engine Skill。
- 已配置华为云访问凭证，建议通过环境变量或安全凭据文件注入，不在对话或文档中暴露AK/SK。
- 目标CCE集群已绑定AOM CCE Prometheus实例。
- 已准备可用的SMN主题，例如主题名test。如果已有AOM通知规则，也可以直接复用。
- 执行账号已具备AOM告警规则查询、创建、通知规则查询和通知规则创建权限。
- 创建告警规则已经过客户二次确认。

涉及的 Skill

Skill名称	功能说明
alarm-correlation-engine	AOM告警查询、活跃告警查询、告警归并分析、告警规则创建和通知规则查询。
huawei-cloud-cce-cluster-management	查询CCE集群列表，确认集群名称、集群ID和集群状态。
observability-context-builder	在告警分析后继续汇聚指标、日志和Kubernetes事件上下文。

操作流程

步骤一：一键创建 CCE 集群告警规则

您可以按照CCE集群告警最佳实践，让AI CLI为目标集群智能配置默认AOM告警规则。您只需要说明区域、集群名称和通知主题，AI CLI会自动理解配置意图，补齐集群信息，生成创建计划，并在客户确认后完成规则创建和结果校验。

在OpenClaw对话框中输入：

帮我给北京四集群test-ai-diagnoses批量创建CCE AOM告警规则，订阅主题使用test。

AI CLI会先生成告警规则配置计划，帮助确认创建范围和通知方式。

智能处理项	说明
识别集群	根据区域和集群名称查询CCE集群，确认集群ID、集群状态和可配置性。
补齐通知配置	优先复用已有AOM通知规则；未指定通知规则时，基于SMN主题创建集群级默认通知规则。
生成规则计划	按CCE集群推荐实践生成50个默认告警规则，区分指标告警和事件告警。
预览影响范围	展示规则数量、规则类型、通知规则和缺失参数，不会直接创建。
等待客户确认	只有客户明确回复确认后，AI CLI才执行创建动作。
自动创建规则	批量创建默认告警规则；如发现同名规则已存在，则自动跳过避免重复创建。
自动验收结果	创建完成后立即查询告警规则列表，确认规则数量、类型和失败数量。

默认创建的50个规则包括。

类型	数量	示例
Prometheus指标告警	38	Pod状态异常、Pod频繁重启、节点CPU使用率、节点磁盘可用率、节点Kubelet异常。
CCE事件告警	12	Pod内存不足OOM、节点磁盘空间不足、节点状态异常、扩容节点超时、集群状态不可用。

AI CLI展示预览后，用户确认规则数量、通知方式和目标集群无误，再明确回复：

确认创建。

执行完成后，AI CLI会返回创建结果，并自动查询确认规则已创建成功。客户重点查看以下验收项：

校验项	期望结果
test-ai-diagnoses相关规则总数	50
指标告警	38
事件告警	12
创建失败	0

如果部分规则已存在，AI CLI会跳过同名规则，避免重复创建。您也可以随时输入以下内容复查：

查询北京四test-ai-diagnoses集群的告警规则。

步骤二：按时间窗聚合分析告警并追查根

当需要了解集群当前是否存在风险时，建议指定时间窗，让AI CLI同时查询活跃告警、历史告警和恢复记录，并按严重程度聚合分析。

分析北京四test-ai-diagnoses集群最近4小时的AOM告警，按严重程度聚合，标记常态告警和突发告警，并告诉我需要先处理的三个问题。

您也可以指定精确时间段，用于分析变更窗口、故障窗口或值班交接窗口。

帮我分析北京四test-ai-diagnoses集群今天09:00到12:00之间的AOM告警，按严重程度聚合，输出突发告警首次出现时间和处置优先级。

AI CLI会先返回本时间窗的告警态势摘要，重点查看以下信息：

输出项	示例
告警总量	最近4小时共发现42条告警，当前未恢复7条
严重程度分布	Critical: 1, High: 3, Medium: 9, Low: 29
告警类型分布	突发告警3组，持续未恢复告警2组，常态高频告警5组
首次出现时间	最早突发告警出现在09:17，集中爆发时间为09:20-09:35
主要影响对象	kube-system、default/nginx-demo、节点192.168.0.12
优先处理建议	当前风险集中在节点资源压力和业务Pod重启，建议优先分析节点磁盘和Pod最近变更

随后，AI CLI会将多条原始告警聚合成告警组，并按严重程度排序。客户可以按以下顺序处理：

优先级	严重程度	告警组	告警特征	首次出现时间	判断	处理建议
P0	Critical	集群状态不可用或核心组件异常	突发告警，当前未恢复	09:17	可能影响集群控制面或业务调度能力	立即进行根因分析，检查CCE事件、核心组件Pod和AOM指标
P1	High	节点磁盘空间不足，关联多个Pod驱逐事件	持续告警，当前未恢复	09:24	节点资源压力可能造成业务副本不稳定	关联节点上的Pod、驱逐事件和磁盘指标
P2	Medium	Pod频繁重启，集中在default/nginx-demo	突发告警，部分已恢复	10:06	可能与近期发布、探针配置或资源限制有关	查询Pod日志、事件和Deployment版本历史
P3	Low	短时CPU阈值告警	常态告警，频繁自动恢复	过去7天多次出现	暂无持续影响，可能是短时流量波动	观察趋势，必要时调整阈值或HPA策略

对于Critical、High或用户关注的告警组，可以继续让AI CLI在同一时间窗内分析相关根因：

继续分析P1节点磁盘空间不足告警，时间窗保持最近4小时，帮我关联这个节点上的Pod、最近事件、相关指标和可能根因。

AI CLI会围绕该告警组继续查询上下文，并给出根因线索：

根因分析项	客户可获得的信息
相关资源	告警节点、受影响Pod、命名空间、工作负载和Service
相关事件	驱逐、调度失败、探针失败、镜像拉取失败、节点异常等事件
相关指标	CPU、内存、磁盘、网络等资源趋势，以及告警触发前后的变化
时间线	告警首次出现时间、集中爆发时间、恢复时间和相关事件发生时间
初步根因	例如节点磁盘压力、工作负载发布异常、核心组件异常或容量不足
下一步建议	继续排障、清理资源、扩容、调整阈值、回滚发布或保持观察

用户也可以要求AI CLI只输出严重且未恢复的告警组：

只看最近4小时仍未恢复的High和Critical告警，按影响资源分组，并给出每组首次出现时间和根因分析建议。

如果告警与业务不可用相关，可以继续要求AI CLI输出完整根因分析和恢复建议：

基于这些严重告警继续做根因分析，时间窗保持09:00到12:00，并给出恢复建议。

如果需要查看原始明细，可以继续输入：

展开最近4小时告警明细，显示告警名称、状态、级别、资源、首次触发时间和描述。

预期结果

完成本实践后，AI CLI能够帮助您完成以下闭环：

- 按CCE集群告警最佳实践，一键为目标集群创建50个AOM告警规则。
- 自动识别集群、准备通知规则、预览创建计划，并在用户确认后执行创建。
- 创建完成后自动查询告警规则列表，确认指标告警38个、事件告警12个、创建失败0个。
- 按指定时间窗查询当前活跃告警、历史告警和恢复记录。
- 对告警进行自动聚合、去重和严重程度标记，区分常态告警、突发告警和持续未恢复告警。
- 输出突发告警首次出现时间、集中爆发时间、影响资源和按优先级排序的处置队列。
- 针对Critical和High严重告警，在同一时间窗内关联事件、日志、指标和资源状态，输出可能根因和下一步诊断路径。

常见问题

提示找不到 Prometheus 实例

- 问题现象
错误示例如下所示。

```
{
  "success": false,
  "error": "未找到目标集群对应的Prometheus实例"
}
```
- 处理建议
 - a. 查询AOM Prometheus实例，确认目标集群是否绑定CCE类型实例。
 - b. 查询CCE插件，确认云原生监控相关组件是否已安装。
 - c. 若控制台刚开启监控，等待实例绑定信息同步后重试。

为什么查询到的通知规则数量和控制台不一致

建议通过AI CLI重新查询当前区域的AOM通知规则；若仍不一致，请核实当前凭证、所属项目及区域是否与控制台环境完全匹配。

为什么自动创建通知规则失效

请优先确认SMN主题是否存在，以及当前账号是否具备AOM通知规则和SMN主题访问权限，确认无误后重新发起创建即可。

告警数量很多但不知道先处理什么

建议通过AI CLI按时间窗重新聚合分析，并要求输出严重程度和处置优先级，例如：

把最近4小时test-ai-diagnoses集群的AOM告警按资源和严重程度聚合，输出最需要优先处理的5个告警组。

如果聚合后仍存在大量High或Critical告警，应继续缩小范围，优先分析未恢复、影响核心命名空间、关联多个资源或集中爆发的告警组。

后续建议

- 规则创建完成后，建议立即查询目标集群的告警规则，确认规则数量、通知方式及启用状态符合预期。
- 新集群上线初期建议每日核查活跃告警与高频历史告警，观察是否存在阈值过严、重复通知或长期未恢复的告警。
- 值班排障时建议固定使用时间窗进行分析。例如，“最近30分钟”“最近4小时”“变更后到现在”，避免无关历史告警干扰。
- 针对Critical和High告警组，建议继续关联日志、事件、指标及工作负载版本，形成根因分析链路。
- 对于确认无业务影响的高频告警，建议先分析触发对象与时段，再评估阈值或通知范围的优化方案。
- 生产集群建议定期导出告警规则清单，将其作为变更审计与故障复盘的重要材料。
- 告警触发后，建议先查看告警聚合和严重程度标记结果，再继续查询相关日志、事件和指标，避免只根据单条告警做恢复动作。

相关文档

- [通过AOM配置自定义告警](#)：了解CCE告警中心中如何通过AOM配置自定义告警，以及Prometheus实例等关键配置项。
- [配置AOM告警规则](#)：了解AOM告警规则的配置方式、规则参数和告警触发逻辑。
- [监控云容器引擎CCE的指标](#)：了解如何使用AOM监控CCE指标，以及指标、事件和告警通知之间的关系。
- [创建主题](#)：自动创建或绑定通知规则前，了解如何在SMN中创建主题。
- [云容器引擎CCE文档](#)：查询CCE集群、插件、云原生观测和工作负载相关产品说明。

2.4 基于 AI CLI 和 Skill 实现 CCE 弹性至 CCI 2.0

CCE突发弹性引擎（对接CCI）是一种虚拟kubelet，将Kubernetes API扩展到无服务器容器平台CCI，支持在短时高负载场景下将CCE工作负载弹性创建到CCI 2.0服务上。本文介绍如何通过AI CLI工具加载cce-cci-bursting-deployer，以一条提示词完成从集群预检到弹性验证的全流程配置。

约束与限制

环境与功能限制

- 仅支持VPC网络模式的CCE Standard集群和CCE Turbo集群。
- 集群所在子网不能与10.247.0.0/16重叠，否则会与CCI命名空间下的Service网段冲突。
- 暂不支持守护进程集（DaemonSet）。
- 安装bursting插件后会在CCI服务新建一个名为“bursting-”+“集群ID”的命名空间，该命名空间完全由bursting插件管理，不建议直接在CCI服务使用该命名空间。

- 具体Kubernetes版本与限制请参见[CCE 突发弹性引擎（对接 CCI）插件功能概览](#)。

IAM 权限要求

API Action	权限	用途
cce:cluster:get	获取集群详情	读取集群网络规格（VPC、子网、ENI）
cce:addon:list	列出插件	检查virtual-kubelet安装状态
cce:addon:create	创建插件	安装virtual-kubelet插件
cce:addon:update	更新插件	配置bursting参数
vpcep:endpoint:create	创建 VPCEP	创建SWR/OBS接口终端节点
vpcep:endpoint:list	列出 VPCEP	检查已有终端节点
vpcep:service:list	列出 VPCEP 服务	发现公共服务详情
vpc:subnet:list	列出子网	验证子网ID
vpc:routetable:list	列出路由表	查找OBS网关路由表 ID

前提条件

- 使用插件前需要在CCI控制台对CCI服务进行授权。
- 如果使用CCI 2.0服务对接CCE突发弹性引擎，请先购买云服务VPCEP，具体操作步骤请参见[环境设置](#)。
- 已安装AI CLI工具并配置华为云凭证环境变量HUAWEI_AK、HUAWEI_SK、HUAWEI_PROJECT_ID。
- 已创建VPC网络模式的CCE Standard集群或CCE Turbo集群，且Kubernetes为1.21及以上版本。
- 已获取OBS VPCEP服务名。OBS终端节点需要精确的obs_endpoint_service_name，该值需要通过华为云工单获取，不能从类似区域的公共服务名推断。如果未提前提供该值，配置流程将在VPCEP创建阶段中断，等待用户补充后才能继续。请在开始配置前通过工单获取该信息并记录，避免流程中断。

操作流程

步骤一：创建 CCE Turbo 集群

创建一个Turbo集群，记录集群 ID 和所在区域。如果已有符合条件的集群，可跳过此步骤。

为我创建一个最新版本的 cce turbo 集群，集群要绑定 eip 用于 kubeconfig 的后续使用。然后为我创建带有两个 4c8g 资源的节点池。

Thinking Done

我来帮你完成这个多步骤任务。首先让我了解相关技能的使用方式。好的，我需要完成以下步骤：创建 CCE Turbo 集群 → 创建节点池 → 绑定 EIP → 获取 kubeconfig → 配置 CCI 弹性。让我先查看集群创建参数的详细文档。开始第一步：检查凭证和 VPC/Subnet 环境。

✔ CCE Turbo 集群创建完成	
项目	详情
集群名称	turbo-cluster-v134
集群 ID	57d85b2f-5dc0-11f1-a40f-0255ac
集群类型	Turbo (ENI 容器网络)
K8s 版本	v1.34
平台版本	cce.3.2
集群规格	cce.s1.small
区域	cn-north-4
公网 API 地址	https://114.116. . . :5443
私网 API 地址	https://192.168. . . :5443
绑定 EIP	114.116.227.199 (5Mbps 按流量计费)

✔ 节点池创建完成	
项目	详情
节点池名称	worker-pool
节点池 ID	a99ba92f-5dc1-11f1-a7f6-0255ac
节点规格	c7.large.2 (4核 8G)
初始节点数	2
操作系统	EulerOS 2.9
系统盘	40GB GPSSD
数据盘	100GB SSD
SSH 密钥	vk-cluster-key
容器运行时	containerd

步骤二：通过 AI CLI 启动弹性配置

cce-cci-bursting-deployer 采用预览优先设计：读操作（precheck、verify、discover、diagnose）可立即执行，写操作（VPCEP 创建、插件安装、工作负载部署）先返回预览方案，用户确认后再执行。

在 AI CLI 中输入以下提示词即可启动全流程：

我的 CCE 集群（集群 ID：xxx，区域：cn-north-4）需要启用到 CCI 2.0 的弹性能力，请按完整流程执行：precheck → VPCEP → 插件安装 → 冒烟部署 → 验证。每个写操作先预览，我确认后再执行。

这个集群，我考虑资源不够的情况下要弹性到 cci，为我准备一下并验证弹性配置是否就绪。

Thinking...

好的，我来为你配置 CCE 到 CCI 的弹性 bursting，并验证配置是否就绪。让我先查看相关技能的文档。好的，工作流程清晰了。让我按步骤执行：Step 1: Precheck 预检 - 检查集群网络、子网角色、addon 状态和阻塞问题：

```
[✓] skill (name=huawei-cloud-cce-cci-bursting-deployer,action=exec)
```

✔ Precheck 通过！关键结果：

- 集群类型：Turbo ✔ (ENI 网络)

Plans

- [✓] Precheck 预检
- [] 节点容量检查
- [] 确保 VPCEP 依赖
- [] Setup CCI Bursting
- [] 验证 Bursting 就绪
- [] 冒烟测试部署与验证

Skill会按以下顺序自动推进，每步涉及写操作时暂停等待确认：

1. 集群预检 (precheck)

调用huawei_precheck_cce_cci_bursting，自动解析集群网络拓扑、区分 cci_subnet_id (Neutron UUID) 和vpcep_subnet_id (VPC UUID) 的子网角色、检查virtual-kubelet插件状态、执行NodeCheck检查物理节点addon headroom。

2. 节点容量检查 (NodeCheck)

如果precheck报告物理节点资源不足，Skill会调用 huawei_check_cce_cci_node_capacity查看详细容量信息，并预览节点池扩容方案，用户确认后执行扩容。

3. VPCEP终端节点创建

调用huawei_ensure_cce_cci_vpcep，自动发现并创建SWR和OBS兼容的接口终端节点。已存在的VPCEP自动复用，不会重复创建。

 注意

OBS终端节点需要精确的obs_endpoint_service_name，请通过华为云工单获取，不要猜测类似区域的公共服务名。如果未提前获取该值，Skill会在此步骤中断并提示用户补充。补充obs_endpoint_service_name后，Skill会继续完成OBS VPCEP终端节点创建，并恢复后续流程。

```
找到了 3 个 OBS 服务 1z026 在 cn-north-4-pek1  
(公共云区)是正确的。但它的端点创建一直失败。SWR/SWR-API VPCEP 已  
accepted, 这是 CCI 镜像拉取的核心依赖。让我直接尝试 setup - 它会自动处理 VPCEP  
并安装 virtual-kubelet addon: Step 4: Setup CCI Bursting - 先预览安装计划:
```

```
[✓] skill (name=huawei-cloud-cce-cci-bursting-deployer,action=exec)
```

```
Canceled
```

```
OBS服务的 vpc ep service名称, 可以使用cn-north-4.com.myhuaweicloud.v4.obs.v2.1z11
```

4. 插件安装

调用huawei_setup_cce_cci_bursting，确认VPCEP依赖就绪后安装或更新virtual-kubelet插件。自动解析并写入区域project ID。该操作幂等：已有插件仅更新配置，不会卸载重装。

```

✔ virtual-kubelet addon 状态为 available - 已就绪! 现在重新运行 setup 来配置
addon 的 CCI 网络参数:

[/] skill (name=huawei-cloud-cce-cci-bursting-deployer,action=exec)

✔ Setup 全部成功! 关键结果:

```

项目	状态
VPCEP	✔ ready=true, missing=[]
OBS v2 端点	✔ accepted
SWR 端点	✔ accepted (IP: 192.168. . .)
SWR-API 端点	✔ accepted (IP: 192.168. . .)
OBS policy	✔ verified
virtual-kubelet 配置	✔ 成功配置 CCI 网络参数

```

已配置参数:
• subnet_id : 7ee78a9d-6d60-4d23-b935-e63221e (CCI Neutron 子网)
• network_id : 同上
• enableScheduleProfileLocalSurge : true
• isInstallProxy : falseStep 5: 验证 Bursting 就绪 - 检查 addon

```

5. 弹性就绪验证

调用huawei_verify_cce_cci_bursting，检查virtual-kubelet插件状态和虚拟节点（bursting-node）是否Ready。如果验证失败，Skill会调用huawei_diagnose_cce_cci_bursting_addon返回结构化诊断报告。

6. 冒烟测试部署

调用huawei_discover_cce_cci_smoke_images发现租户自有SWR基础版镜像，然后调用huawei_deploy_cce_cci_smoke_workload创建Deployment。该Deployment自动添加bursting.cci.io/burst-to-cci: enforce标签强制调度到CCI，不指定image参数时自动选用发现的租户镜像。

```

🚀 冒烟测试成功!
两个 Pod 都在 bursting-node 上 Running, 已成功弹性到 CCI:

```

Pod	状态	IP	节点
cci-smoke-nginx-cc4b5b7b4-684n7	1/1 Running	192.168. . . 0	bursting-node (CCI)
cci-smoke-nginx-cc4b5b7b4-6999j	1/1 Running	192.168. . .	bursting-node (CCI)

7. 最终验证

再次调用huawei_verify_cce_cci_bursting，确认所有Pod在CCI虚拟节点上达到Running状态，并在CCE控制台可见。

```

✔ CCE>CCI 弹性 bursting 配置与验证全部完成!

最终状态总结

```

资源	状态
CCE Turbo 集群	turbo-cluster-v134 , v1.34, 4 物理节点 Ready
EIP	114.116.227.199 (公网 kubeconfig 访问)
节点池	worker-pool , 4x c7.large.2 (4c8g), 全部 Ready
virtual-kubelet addon	v1.5.82, available ✔
bursting-node 虚拟节点	Ready ✔
冒烟测试 Pod	2/2 Running on bursting-node (CCI) ✔
VPCEP	SWR/SWR-API/OBS 全部 accepted ✔

```

冒烟测试结果: nginx:latest 的 2 个副本成功调度到 bursting-node (CCI 虚拟节点), Pod 获得 CCI Neutron 子网 IP (192.168.3.x), 状态 Running, CCE>CCI 弹性通道已就绪.

```

步骤三：配置 ScheduleProfile 调度策略

弹性基础配置完成后，可通过创建ScheduleProfile控制工作负载调度行为。

调度策略	说明	适用场景
localPrefer	优先调度到CCE，资源不足时弹性到CCI	日常弹性扩容
enforce	强制调度到CCI	测试验证、CI/CD临时任务
auto	由调度器打分决定是否弹性到CCI	灵活调度

在AI CLI中输入以下提示词创建ScheduleProfile：

请帮我创建一个 ScheduleProfile，命名空间 default，
匹配标签 app=nginx，策略为 localPrefer，
本地最大实例数 20，CCI 缩容优先级 10。

ScheduleProfile 创建成功 现在创建匹配该规则的 nginx Deployment：

📖 说明

label策略优先级高于ScheduleProfile。如果Pod同时有bursting.cci.io/burst-to-cci: off标签和enforce profile，Pod不会被调度到CCI。

至此，CCE到CCI 2.0的弹性配置与调度策略全部完成。

ScheduleProfile 创建结果		
字段	值	
名称	nginx-local-prefer	
命名空间	default	
匹配标签	app: nginx	
策略	localPrefer (本地优先)	
本地最大实例数	20	
本地缩容优先级	0	
CCI 缩容优先级	10	
虚拟节点	bursting-node	
状态	initialized <input checked="" type="checkbox"/>	
nginx Deployment 分布结果 (25 副本 → localPrefer 策略生效)		
位置	Pod 数量	说明
本地物理节点	10 个	local.maxNum=20, 当前用了 10 溢出的副本自动弹出到 CCI
bursting-node (CCI)	15 个	
总计	25/25 Running <input checked="" type="checkbox"/>	

常见问题与诊断方式

当弹性配置过程中遇到问题，可在AI CLI中直接描述现象，Skill会调用相应诊断工具返回结构化报告。

问题现象	诊断提示词
CCI Pod ImagePullBackOff或镜像拉取超时	CCI Pod拉镜像失败，请诊断
虚拟节点不Ready	虚拟节点不Ready，请诊断addon
addon Pod Pending或反复重启	addon Pod Pending，帮我检查节点容量
addon日志报region mismatch	addon日志报region不匹配，请诊断
addon日志报IAM denied或project ID missing	addon报IAM denied，请诊断
OBS网关节点failed，VPCEP创建失败	OBS VPCEP创建失败，请使用我通过华为云工单获取的服务名：<实际的obs_endpoint_service_name>重试

相关文档

- [CCE容器实例弹性伸缩到CCI服务](#)
- [CCE突发弹性引擎（对接 CCI）插件功能概览](#)
- [CCI 2.0快速使用](#)
- [CCI 2.0环境设置](#)

2.5 基于 Hermes 与飞书构建 CCE 生产环境智能运维 Agent

本文以Hermes接入飞书为例，构建一个CCE生产环境ChatOps值班Agent。该Agent可以定时扫描现网告警，自动归并和分析告警，生成恢复方案，并在用户通过飞书手机端确认后执行恢复动作。文中的CPU高告警只是一个用于验证闭环的小案例，您可以基于同样的思路扩展出Pod重启诊断、节点异常处理、调度失败恢复、容量巡检、发布变更关联、日报生成等自定义能力。

应用场景

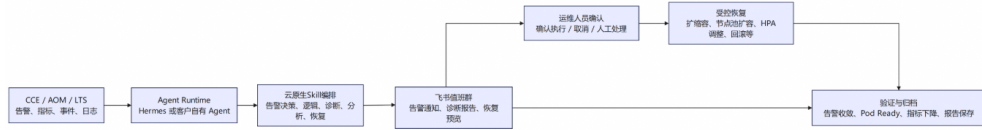
在生产环境中，CCE集群可能持续产生大量告警，覆盖工作负载、Pod、节点、网络、存储、弹性伸缩、资源容量和可用性风险等多个维度。现网规模越大，告警数量、告警来源和处理路径越复杂，值班人员需要在飞书告警、AOM指标、Kubernetes事件、Pod日志、工作负载配置、HPA状态、节点容量和工单系统之间反复切换，容易出现告警疲劳、响应慢、证据不完整、恢复动作缺少审核和复盘材料难沉淀等问题。

通过构建CCE生产环境智能运维Agent，您可以将“告警发现、告警归并、上下文采集、根因分析、恢复预览、用户确认、执行恢复、效果验证、结果归档”沉淀为一套可复用的ChatOps值班能力。Agent可以使用Hermes，也可以使用您已有的ChatOps、AIOps、工单助手或自研值班机器人；华为云云原生Skill负责提供CCE、AOM、LTS、节点池、工作负载等资源的标准化查询、分析和受控恢复能力。

方案架构

本方案采用“Agent Runtime + 云原生Skill + 飞书确认”的组合架构。Agent Runtime负责任务调度、告警分发、上下文编排和飞书交互；云原生Skill提供告警、指

标、日志、事件、根因分析和恢复动作能力；飞书承载告警通知、用户审核和闭环结果。



建议按能力边界拆分Agent权限。

能力域	典型动作	推荐控制方式
告警治理	查询、归并、分级、路由AOM 活跃告警和历史告警	只读权限，允许定时自动执行
诊断分析	汇聚指标、事件、日志、工作 负载、节点状态	只读权限，允许Agent自动编排
恢复预览	生成扩容、回滚、HPA、节点 池等恢复方案	只生成预览，不修改资源
恢复执行	执行变更动作	必须经过飞书确认
审计归档	保存告警、分析、确认、执行 和验证记录	建议进入工单、OBS或内部知识库



约束与限制

- 告警规则和指标口径应使用云服务推荐配置，并结合实时指标做交叉验证。
- 恢复类动作必须保留预览和确认机制，尤其是扩缩容、回滚、节点池变更等生产操作。
- 飞书消息应面向值班人员可读，优先输出结论、证据、影响对象、可选方案和确认入口。
- 对复杂恢复链路，建议将告警指纹、方案编号、目标资源和执行记录持久化，便于确认和审计。

- 禁止在提示词、飞书消息、截图或文档中暴露AK/SK、Token、证书、Project ID等敏感信息。
- 在试点阶段，建议从只读巡检和人工确认式恢复开始，再逐步扩展到更自动化的恢复策略。

前提条件

执行本实践前，建议先准备好Agent运行环境和CCE观测对象，后续再逐步扩展自动化范围：

- 已准备用于巡检和诊断的CCE集群、命名空间或业务范围。
- 已接入AOM告警、云原生监控指标或已有巡检对象。
- 已准备Hermes或自有Agent Runtime，并接入飞书或工单系统。
- 已将CCE相关华为云云原生Skill导入Agent，用于查询告警、指标、事件、工作负载、Pod、节点和执行受控恢复。
- 已准备好访问凭证，且通过安全方式配置，确保不在提示词或文档中暴露敏感信息。

可编排能力

您可按需组合以下Skill，构建面向海量告警的智能运维流程。

能力	代表Skill	作用
告警发现与归并	alarm-correlation-engine	查询活跃/历史告警，归并重复告警，识别需要处理的告警。
可观测上下文	observability-context-builder	汇聚指标、日志、事件和资源状态。
指标分析	metric-analyzer	分析CPU、内存、网络、磁盘等趋势。
Pod 诊断	pod-failure-diagnoser	分析Pod状态、重启、日志和Events。
工作负载诊断	workload-failure-diagnoser	分析Deployment、ReplicaSet、HPA、Service和Endpoint。
节点诊断	node-failure-diagnoser	分析节点状态、资源水位和调度能力。
变更关联	change-impact-analyzer	关联告警前后的发布、配置和资源变更。
根因分析	root-cause-analyzer	汇总证据，输出根因、置信度和建议。
受控恢复	auto-remediation-runner	生成恢复预览，确认后执行恢复动作。

面向现网告警治理时，可将Agent能力分为以下层次。

层次	目标	示例
告警入口	接收和发现不同来源的告警	AOM告警、巡检任务、飞书消息、工单事件
告警治理	降低告警噪声并确定处理优先级	去重、归并、分级、路由、静默、摘要
智能诊断	从多源数据中定位候选原因	告警、指标、日志、事件、变更、资源状态
受控恢复	将建议动作转为可审核的恢复方案	扩缩容、HPA 调整、节点池扩容、回滚
闭环运营	将处理结果沉淀为可复用经验	飞书通知、工单归档、日报、复盘材料

操作流程

本实践基于集群demo-recovery中default/chat-app工作负载的CPU使用率过高告警展开，旨在验证CCE生产环境智能运维Agent的端到端能力。CPU高告警只是现网告警中的一种类型，流程重点是展示从巡检正常态、告警发现、自动分析、用户确认、受控恢复到闭环验证的完整链路。

步骤一：启动巡检机器人

初始化Agent后，加载CCE相关Skill，并配置巡检周期和飞书通知目标。

巡检机器人应能够在“无告警”和“有告警”两种状态下都输出清晰结果，避免值班人员无法判断巡检链路是否正常。



正常会输出“集群demo-recovery当前无异常, 环境正常”。上图展示了巡检机器人在未发现活跃告警时, 通过飞书输出集群健康摘要。客户可以将该能力扩展为每日巡检、班前巡检或重点业务巡检。

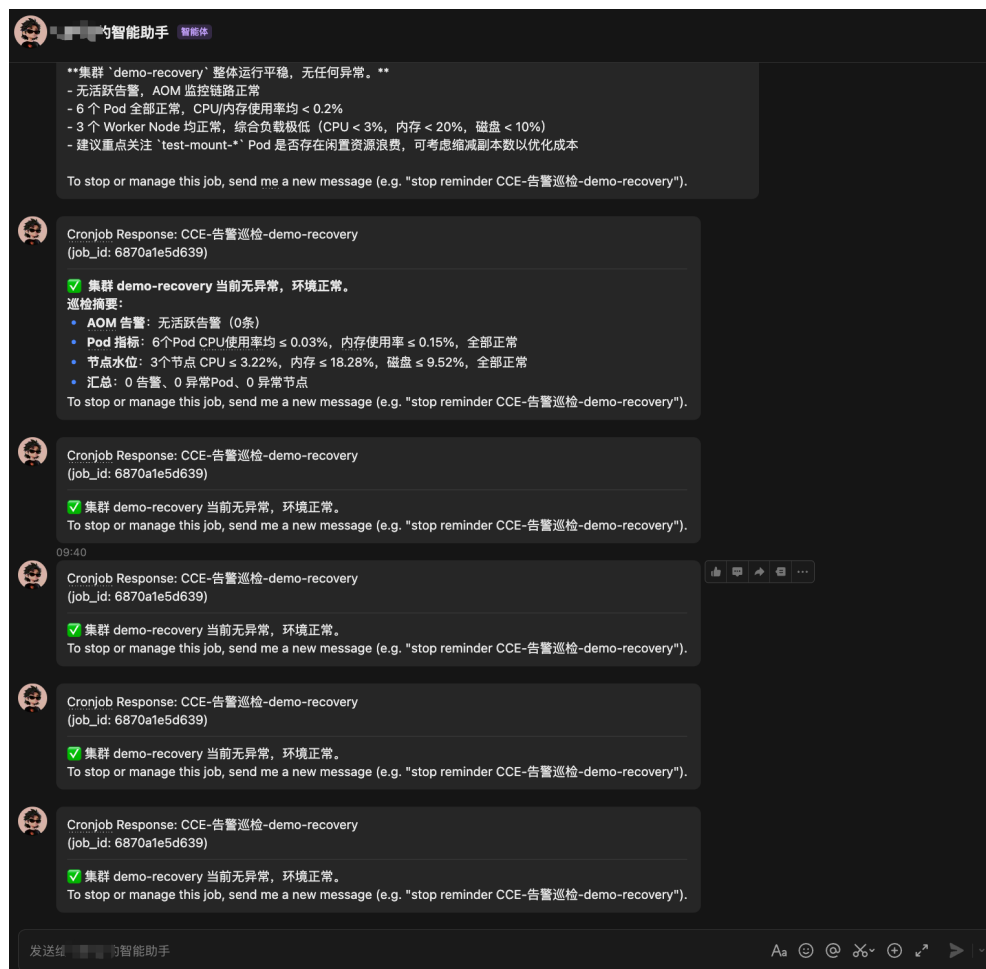
步骤二: 接收 CPU 高告警

当AOM产生CPU高告警后, 巡检机器人在飞书中输出告警摘要, 并将该告警纳入自动分析流程。生产环境中, 同一入口也可以接收Pod重启、节点异常、调度失败、Service无后端、HPA不生效等其他类型告警。

收到告警后, Agent应同时关注AOM告警状态和实时指标, 避免只依据单一信号做判断。

步骤三：自动分析告警并生成恢复预览

巡检机器人发现CPU高告警后，自动收集告警、Pod指标、节点水位、工作负载状态和影响对象，形成诊断报告。报告中应突出可观测事实、证据链和可选恢复方案。



上图展示了巡检机器人在未发现活跃告警时，通过飞书输出集群健康摘要。客户可以将该能力扩展为每日巡检、班前巡检或重点业务巡检。

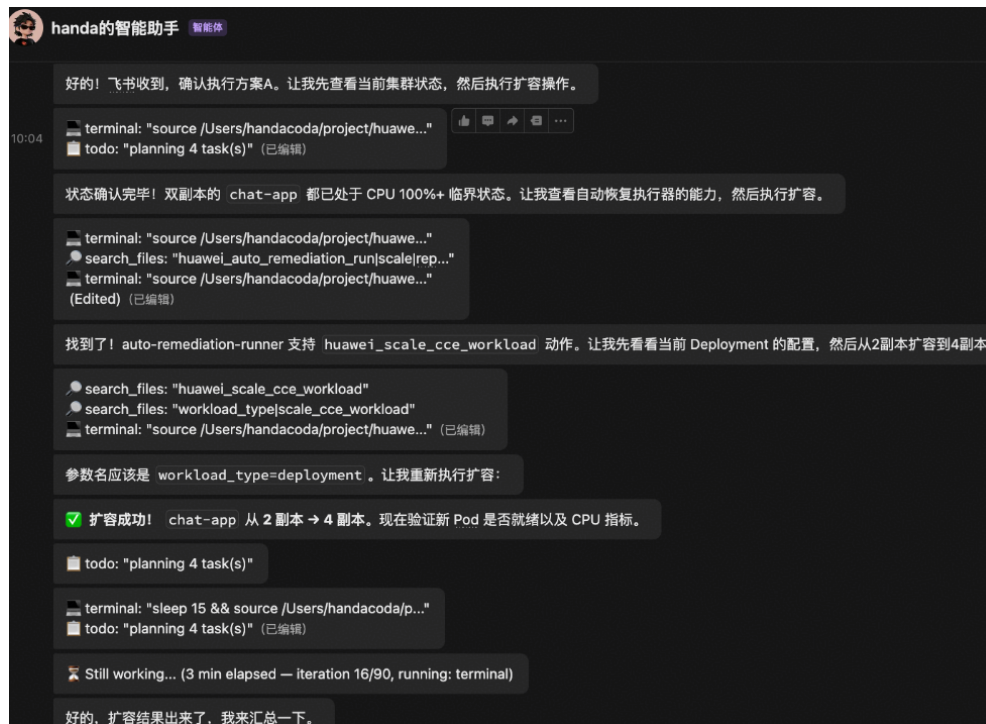
恢复预览建议包含以下项目。

项目	内容
告警摘要	告警名称、级别、状态、触发时间
影响对象	集群、命名空间、工作负载、Pod、节点
关键证据	CPU水位、节点水位、Pod状态、相关告警
候选方案	扩容副本、调整资源、节点池扩容、人工处理等
变更影响	资源占用、调度条件、成本变化、回滚方式
用户确认	明确给出确认语句或按钮

步骤四：在飞书确认恢复方案

这是生产变更链路中的关键审核环节。Agent在此阶段仅等待用户确认，不会执行扩缩容、回滚、重启、节点池变更等任何改变现网状态的操作。只有当用户在飞书中明确回复“确认执行”或“确认执行方案A/B”等确认语句后，Agent才会按已确认的方案执行恢复操作。

确认语句应与恢复预览中的方案编号保持一致，例如“确认执行方案A”对应扩容副本，“确认执行方案B”对应资源升配或容量扩展。您也可以将确认动作替换为飞书卡片按钮、工单审批或企业审批流，但核心要求保持不变，即未经人工确认，Agent仅做分析和预览，不进行现网变更。



步骤五：执行恢复并持续复核

本案例首先选择扩容工作负载，将chat-app从2个副本扩容到4个副本。扩容后，Agent继续复核Pod状态和节点容量，发现其中一个新增Pod因节点CPU资源不足而Pending。

```
| 回复 handa: 确认执行方案A
**方案A执行完成! ** 以下是完整闭环报告:

---

## ✅ CCE 恢复完成报告 — demo-recovery

### 执行动作
| 项目 | 内容 |
|-----|-----|
| **动作** | 扩容副本 |
| **目标** | Deployment `chat-app` (namespace: default) |
| **变更** | 副本数 **2 → 4** |
| **状态** | ✅ 已完成 |

### 验证结果
| Pod | 状态 | CPU | 节点 |
|-----|-----|-----|-----|
| `chat-app-c4c89c544-7bhnd` | ✅ Running | **99.1%** (critical) | 192.168.6.208 |
| `chat-app-c4c89c544-jv4p9` | ✅ Running | **102%** (critical) | 192.168.3.233 |
| `chat-app-c4c89c544-w6gtp` | ✅ Running (新) | 刚启动, 指标未采集 | 192.168.3.233 |
| `chat-app-c4c89c544-fxq7` | ❌ Pending** | — | Unschedulable (CPU不足) |

### 注意事项
- **新 Pod `w6gtp`** 已成功启动并运行, 正在分摊 CPU 负载
- **新 Pod `fxq7`** 因 **3个节点 CPU 资源不足** 无法调度 (每个 Pod 请求 1 核, 当前节点无足够空闲资源)
- 旧的两个 Pod 仍处于 ~100% CPU 临界状态
- AOM 告警规则使用 raw counter 而非 `rate()`, 告警存在振荡 (已知问题)

### 后续建议
当前扩容到 4 副本已尽力利用节点资源。如果需要让第 4 个 Pod 也能运行, 可以考虑:
1. **方案B: 升配节点** — 扩容节点池增加节点资源
2. **等待下次巡检自动验证** — cron 每 5 分钟会检查 CPU 是否降到 80% 以下

---
```

这个分支体现了生产恢复中很重要的一点, 即恢复动作需要持续验证。扩容请求成功并不代表所有 Pod 已调度成功, 也不代表告警已经收敛。Agent 应将复核结果继续反馈给用户, 并给出下一步方案。

步骤六: 追加容量动作并完成闭环

当扩容受节点容量限制时, Agent 可以生成新的容量恢复方案, 例如节点池扩容、调整工作负载 request、优化 HPA 上限或接入 CCI 弹性能力。本案例选择追加节点池扩容, 等待新节点上线后由 Kubernetes 自动调度 Pending Pod。



执行恢复动作后，Agent再次巡检告警、Pod、节点和CPU指标，并将闭环结果发送到飞书。



Hermes 任务提示词参考

以下提示词可作为Hermes ChatOps值班Agent的任务模板。它弱化了具体命令和环境变量，只保留角色、流程、输出结构和安全边界。

您是CCE生产环境智能运维Agent，负责对目标CCE环境执行告警巡检、告警归并、自动分析、恢复预览、用户确认后恢复、恢复验证和飞书闭环通知。

前置约定：

- CCE相关云原生Skill已提前导入当前Agent。
- 目标集群、巡检范围、通知通道和访问凭证已由运行环境提供。
- 所有通知都发送到飞书或客户指定的值班通道。

目标：

1. 定时扫描目标CCE环境的活跃告警和近期历史告警。
2. 对告警进行去重、归并、分级、路由和影响范围摘要。

3. 巡检正常时，输出简洁的健康摘要，不静默退出。
4. 发现需要处理的告警时，自动汇聚上下文，包括AOM告警、实时指标、Kubernetes Events、Pod/Workload/Node 状态、日志摘要和近期变更。
5. 输出面向值班人员的诊断报告，包含告警摘要、影响对象、关键证据、候选原因、建议方案和需要确认的动作。
6. 对涉及资源变更的动作，只生成恢复预览，不直接执行任何现网动作。
7. 只有用户在飞书中明确回复“确认执行”或确认具体方案后，才执行恢复动作。
8. 执行后必须验证告警状态、Pod状态、工作负载副本、节点容量和关键指标，并将闭环结果发送到飞书。

巡检报告建议结构：

- 巡检摘要：集群、时间窗、活跃告警数量、关键资源状态。
- 告警发现：告警名称、级别、状态、影响对象、当前观测值。
- 系统分析：告警状态、实时指标、Pod/Workload/Node 状态、相关事件和近期变更。
- 恢复方案：给出2到3个可选方案，说明适用场景、影响范围、回滚方式和验证方式。
- 用户确认：明确提示用户回复“确认执行”或选择具体方案。

安全边界：

- 告警扫描、证据采集和根因分析阶段只允许只读操作。
- 不允许仅凭单条告警直接执行恢复。
- 所有写操作必须先输出恢复预览、影响范围、回滚方式和验证方式。
- 同一条告警中的方案编号和方案含义保持一致，用户确认后按已确认方案执行。
- 在用户确认前，禁止执行扩缩容、回滚、重启、节点池变更等任何会改变现网状态的动作。
- 不在输出中暴露AK/SK、Token、证书、Project ID等敏感信息。
- 如果证据不足，列出可能原因和需要人工复核的信息，不替用户做没有证据支撑的判断。

飞书输出要求：

- 巡检正常时，输出简洁的健康摘要。
- 发现告警时，先输出告警摘要和影响对象，再输出关键证据和候选方案。
- 需要恢复时，明确提示用户回复“确认执行”或选择具体方案；在收到人工确认前，只等待确认，不执行变更。
- 恢复完成后，输出执行动作、验证结果、剩余风险和后续建议。

诊断结果

这个案例的重点不在于CPU高一定要扩容，而在于展示一条可迁移的方法：Agent发现问题 → Skill汇聚证据 → 人工确认恢复 → 系统执行并验证。客户可以将其中任何一段替换成自己的工具、审批和业务规则。

阶段	结果
告警发现	检测到CCE容器CPU使用率大于80%
自动分析	汇聚AOM告警、Pod指标、节点水位和工作负载状态
恢复预览	给出扩容副本等可选方案，并等待飞书确认
首次恢复	将工作负载从2副本扩容到4副本
过程复核	发现一个新增Pod因节点资源不足Pending
追加动作	追加节点池扩容，补充调度容量
闭环验证	活跃告警清零，Pod和节点巡检正常，飞书输出结果

扩展应用场景

您可以根据自身需求和现有工具，从多个维度扩展本实践，以适应不同的运维场景和业务需求。以下是从不同维度扩展本实践的具体建议和示例。

扩展方向	示例
更换Agent	使用Hermes、OpenClaw、AI CLI、企业ChatOps机器人或自研Agent。不同的Agent可以提供不同的功能和集成能力，您可以根据自己的技术栈和需求选择合适的Agent。
更换入口	从AOM告警、飞书消息、工单、定时任务、发布事件或人工问询触发。不同的入口可以适应不同的告警来源和触发方式，提高告警处理的灵活性和响应速度。
更换Skill组合	针对Pod、Node、Network、Storage、HPA、Cost等场景编排不同Skill。通过组合不同的Skill，您可以针对特定的运维场景提供更精准的诊断和恢复能力。
更换审批方式	使用飞书回复、飞书卡片按钮、工单审批、变更审批流。不同的审批方式可以适应不同的企业审批流程和安全要求，确保恢复动作的合规性和安全性。
更换恢复动作	扩缩容、HPA 调整、节点池扩容、回滚、隔离节点、停止异常任务。不同的恢复动作可以应对不同的故障类型和恢复需求，提高恢复的灵活性和有效性。
更换归档方式	输出到飞书、工单、OBS、日报、知识库或审计系统。不同的归档方式可以满足不同的记录和审计需求，确保告警处理过程的可追溯性和可复盘性。

典型扩展场景如下表所示。

场景	编排思路
Pod频繁重启	汇聚重启次数、previous日志、OOM、探针配置和Events，生成回滚或资源调整预览。通过分析Pod的重启次数、日志、OOM事件、探针配置和相关事件，生成回滚或资源调整的恢复方案，帮助客户快速定位和解决 Pod 频繁重启的问题。
Pod Pending	分析节点容量、污点容忍、亲和性、PVC、镜像拉取和配额，生成调度恢复建议。通过分析节点的资源水位、污点容忍、亲和性、PVC状态、镜像拉取和配额限制，生成调度恢复建议，帮助客户解决Pod无法调度的问题。
节点异常	关关节点状态、资源水位、组件状态和事件，生成隔离、迁移或节点池扩容预览。通过分析节点的状态、资源水位、组件状态和相关事件，生成隔离、迁移或节点池扩容的恢复方案，帮助客户快速处理节点异常问题。
Service无后端	分析Deployment、Endpoint、Service Selector和发布状态，定位发布或选择器问题。通过分析Deployment、Endpoint、Service Selector和发布状态，帮助用户定位Service无后端的原因，生成相应的恢复建议。
HPA不生效	分析指标采集、request配置、HPA上下限和扩缩容事件。通过分析指标采集、request配置、HPA上下限和扩缩容事件，帮助客户诊断HPA不生效的原因，生成相应的恢复建议。

场景	编排思路
定期巡检	定时输出告警、资源水位、异常Pod、节点风险和成本优化建议。通过定时巡检，输出告警、资源水位、异常Pod、节点风险和成本优化建议，帮助客户及时发现和处理潜在问题，优化资源使用。

预期结果

完成本实践后，用户可以获得以下效果：

1. CCE告警进入飞书后，Agent自动启动诊断链路。
2. 运维人员无需在多个系统间反复切换，即可看到告警摘要、证据和候选方案。
3. 恢复动作在执行前经过飞书确认，降低误操作风险。
4. 恢复执行后自动验证告警收敛、Pod状态、节点容量和指标趋势。
5. 告警处理过程可归档、可审计、可复盘。
6. 同一套Agent + Skill编排思路可扩展到更多CCE运维场景。