

ZENTRIX GATEWAY

企业 AI 产品文档

配置 · 使用 · 结果核对

Zentrix Gateway 管理控制台操作指南

配置组织、模型和工具，验证权限与流量规则，找到运行问题的处理入口。

操作与使用 | 权限与边界 | 结果核对

文档版本

2026.10

核对日期

2026 年 10 月 14 日

目录

1. 进入控制台与菜单可见性

2. 当前导航速查

3. 建立组织与权限

4. 接入模型并开放模型服务

5. API 密钥与调用身份

6. 接入 MCP、API 服务与外部能力

7. 护栏与内容检测

8. 流量规则、网关扩展与域名

9. 用量、观测与排障

10. 告警、导出与审计

11. 平台维护（平台管理员）

12. 首次上线验收

> 2026.10 · 适用于 Web 管理控制台；可见菜单由管理范围、角色和许可模块决定。各项操作的完整步骤见《Zentrix Gateway 用户手册》。

> 页面名称按当前源码核对；角色、许可和部署决定实际可用范围。正式操作前，以所在环境显示的配置与状态为准。

1. 进入控制台与菜单可见性

从能力中心右上角账号菜单选择「管理控制台」。侧边栏顶部显示「管理后台」和当前租户名称，「返回工作台」回到能力中心「首页」；内容区顶部以「分组 › 菜单」显示当前位置。

菜单由三个条件共同决定，看不到某个菜单时按以下顺序核对：

1. 管理范围：租户管理员使用本租户的管理控制台；平台管理员（安装时创建的 root，或在「平台账号」中被授予「平台管理员」的平台账号）使用平台范围的管理控制台，进入后默认打开「租户管理」。两种控制台显示的菜单不同，见第 2 节。
2. 角色：租户属主默认拥有全部权限，其余职责通过「角色管理」中的六个内置角色分配。

角色	职责	侧边栏可见菜单
租户管理员	本租户的全部配置	全部租户菜单，包括只对租户管理员显示的「会话诊断」「角色管理」
系统管理员	配置与装配，不触及凭据与审计	「概览」、「搭建」分组、「成员与部门」「职能组」「网关扩展」「报表导出」、「系统设置」分组（「全局设置」中显示「常规」「品牌与外观」「授权许可」页签）
安全管理员	凭据与安全策略	「SSO 配置」「护栏」「流量规则」；「全局设置」中只显示「账号与会话安全」页签
审计员	审计日志与数据留存	「操作审计」；「全局设置」中只显示「数据留存」页签
运营	值班运营与告警处置	「告警规则」
财务、只读审计	只读查看费用，或只读查看请求与对话明细	无侧边栏菜单，由管理员提供「概览」页面地址后直接打开

   系统管理员、安全管理员、审计员两两互斥，同一账号只能持有其中一个；运营、财务、只读审计可与任一角色叠加。角色只能授予单个用户。

3. 许可模块：许可证未包含某个加购模块时，对应菜单不显示，相关接口拒绝访问。

许可模块	对应界面
基础功能	下列加购模块以外的全部菜单
模型护栏	「护栏」：敏感数据、词库、问答、内部系统白名单、威胁情报
动态路由	「流量规则」：规则清单、全局限额、意图集、缓存与验证规则

许可模块	对应界面
全链路审计	「操作审计」
FinOps	「概览」中的用量、费用与耗时面板；「全局设置 → 「授权许可」页签」中显示为「用量与成本」

「知识库管理」另需许可证包含知识库能力且部署已安装知识库。角色与菜单、「概览」各页签的可见角色见《Zentrix Gateway 用户手册》第 3.4 节与附录 A。

2. 当前导航速查

租户管理控制台

分组	菜单	页签或主要用途	手册章节
概览	概览	「网关总览」「网关请求」「对话」「路由策略命中」「网关告警」	第 17 章至第 20 章、第 22 章
概览	会话诊断	自动识别会话中的问题，给出原因、证据与处理建议	第 20 章
搭建	插件、技能、外部智能体、外部注册中心	导入与分发技能和插件包；接入外部智能体（「Agent 目录」「任务中心」）；从 Nacos 同步	第 12 章
搭建	MCP 服务、API 服务	「MCP 列表」「Token 管理」「分类管理」；「API 服务列表」「Token 管理」「分类管理」	第 11 章
搭建	知识库管理	维护知识源与检索资产	—
搭建	模型服务	「模型服务」「API 密钥管理」	第 6 章、第 7 章
搭建	模型供应商	「模型供应商」「模型&定价」「分类管理」	第 4 章、第 5 章
组织与权限	SSO 配置、成员与部门、角色管理、职能组	身份源、部门与成员、角色、跨部门成员集合	第 3 章
护栏	护栏	「敏感数据」「词库」「问答」「内部系统白名单」「威胁情报」	第 13 章、第 14 章

分组	菜单	页签或主要用途	手册章节
路由策略	流量规则	「规则清单」「全局限额」「意图集」「缓存」，另有「验证规则」页面	第 15 章
路由策略	网关扩展	浏览、上传、安装和绑定 WASM 网关扩展	第 16 章
运行与排障	告警规则、SMTP 配置	「告警规则」「通知渠道」；系统邮件	第 22 章
审计与合规	操作审计、报表导出	配置变更与访问记录、审计证据；调用与用量数据导出	第 21 章、第 22 章
系统设置	通知横幅、全局设置	本租户公告；全局设置按页签划分：「常规」（各功能的默认模型与美元汇率）、「品牌与外观」（产品显示名、标志、网站图标、主题色）、「账号与会话安全」（登录会话与「安全策略」）、「数据留存」（内容留存与留存时长）、「授权许可」（许可证）	第 5.10 节、第 5.7 节、第 24.8 节、第 3.7 节、第 19.4 节、第 24.7 节

平台范围的管理控制台

分组	菜单	主要用途
租户	租户管理、通知横幅	租户生命周期；对所有租户生效的公告
网关与网络	网关、出口网络、SSRF 出站防护	公开访问地址、入口证书、网关节点与过载保护；平台访问公网的质量；出站访问防护
运行与排障	告警规则、系统运行监控	「活跃告警」「规则目录」「通知渠道」；平台各组成部分的运行状态
平台账号与登录	平台账号、SSO 配置、全局设置	平台管理人员的个人账号；平台单点登录（OpenID Connect）；全局设置只有「账号与会话安全」页签（登录会话、「安全策略」与「管理面访问网段」）
系统维护	系统更新、备份与恢复、部署配置	升级；备份与恢复；部署配置（含资源容量）与入口监听

常见任务的入口

要做的事	入口
查看全租户用量与费用	「概览 → 网关总览」（需要 FinOps）
查询一次调用的结果、路由与上游尝试	「概览 → 网关请求」，打开请求详情
查看规则命中	「概览 → 路由策略命中」；限额与护栏的命中在请求详情的「命中记录」中
处置告警事件	「概览 → 网关告警」
模拟一次请求的规则判定	「流量规则 → 规则清单」右上方的「验证规则」
设置请求次数、并发与熔断	「流量规则 → 全局限额」
检索、轮换、禁用 API 密钥	「模型服务 → API 密钥管理」
设置 Key 轮换宽限期与 令牌有效期	「系统设置 → 全局设置 →「账号与会话安全」页签 → 安全策略」
指定内容安全、敏感信息识别、意图识别等默认模型	「系统设置 → 全局设置 →「常规」页签」

3. 建立组织与权限

1. 在「成员与部门」中新建部门、添加成员。「添加成员」中开启「管理员权限」的成员成为租户管理员，只授予确实需要管理整个租户的人员。
2. 在「职能组」中建立跨部门的成员集合，用于分发模型服务等能力。
3. 在「角色管理」中为系统管理员、安全管理员、审计员等角色指定不同人员，用各角色账号登录核对侧边栏。
4. 需要企业身份登录或通讯录同步时，在「SSO 配置」中「添加身份源」：钉钉、飞书、企业微信、企业 IAM (OAuth2)、LDAP 或目录推送。完成测试登录或连接自检后再启用。
5. 在「系统设置 → 全局设置 →「账号与会话安全」页签」中选择「安全策略」，设置会话超时、在线人数、登录方式、强制两步验证、密码策略与外部访问的 令牌有效期。

部门与职能组决定成员能使用哪些能力，角色决定成员能进入哪些管理菜单。成员离职时选择「删除」，账号移入「已离职成员」，其 API 密钥与登录凭据立即失效。详见《Zentrix Gateway 用户手册》第 3 章。

4. 接入模型并开放模型服务

1. 在「模型供应商」中选择「添加模型供应商」，选择供应商类型，核对「请求地址」与「模型请求格式」，填写 API 密钥或订阅账号，选择「验证并获取可用模型」，启用需要的模型后保存。核对可用模型时，页面内常驻展示连通测试结果：多网关节点部署下，测试会按节点逐一探测并展示各节点的连通状态、耗时与失败层级；模型列表明确标示来自「供应商接口实时获取」还是「预置目录回退」；若部分节点暂不可达但需先行保存草稿，

可勾选页面显示的「仍要保存」确认项。在供应商详情页的「探活记录」中可按网关节点筛选历史探活结果，排查特定节点的网络、代理或证书问题。

2. 在「模型&定价」中为需要计费的模型设置售价，可用单个计价、阶梯计价、分时计价或参照目录价。未定价的模型仍可调用，但不计入金额，且会被金额预算拒绝。
3. 在「模型服务」中选择「添加模型服务」，加入至少一个已启用的模型。新服务的「可见范围」显示“未分发”。
4. 在服务详情的「试跑」页签发起真实调用，确认成功；失败时按页面给出的失败位置与原因处理。
5. 在「可用范围与配额」中选择「编辑可用范围」，先授权试点部门、职能组或成员，并设置有效期；再按对象「配额编辑」设置 Token 预算与金额预算（小时、天、月、累计）。
6. 在「系统设置 → 全局设置」中为内容安全、敏感信息识别、意图识别和问答语义匹配指定默认模型，所选模型需要已声明对应的能力用途。



配额只限制当前模型服务；单次请求 Token 上限、并发和每分钟请求数由流量规则与全局限额控制。平台不会因为某个模型不可用而自动改用服务内的其他模型，需要故障转移时在「流量规则」中配置。详见《Zentrix Gateway 用户手册》第 4 章至第 7 章。

5. API 密钥与调用身份

API 密钥由成员在能力中心创建，每把 Key 绑定一个模型服务；管理员不在管理控制台新建 Key。管理员在「模型服务 → API 密钥管理」中：

- 按名称或前缀、状态、模型服务、调用方检索全部 Key；
- 「轮换密钥」：新密钥只显示一次，旧密钥在「安全策略」设置的宽限期内仍可用（默认 24 小时）；
- 「禁用」或「删除」：调用立即被拒绝，禁用后不能在此重新启用；发现泄露时使用这两项，不使用轮换；
- 「编辑配额」：为单把 Key 设置独立额度池。

业务系统也可以携带企业身份令牌直接调用模型：在「SSO 配置」的企业 IAM (OAuth2) 身份源中开启「允许业务系统用身份源令牌调用模型」。调用身份只来自平台认证结果，普通请求头中声明的身份不构成授权。详见《Zentrix Gateway 用户手册》第 6.7 节、第 9 章。

6. 接入 MCP、API 服务与外部能力

对象	关键步骤	手册章节
MCP 服务	「添加应用」→ 选择接入方式（远程 MCP URL、MCP Json、OpenAPI、Postman Collection、cURL 或手动添加）→ 解析并核对工具 → 配置「MCP 鉴权」→ 打开「发布」→「编辑可用范围」；工具级访问控制、参数匹配规则和请求体积上限在「路由策略 → 流量规则」中配置；MCP 详情「基本信息」中的「查看适用的路由策略」直接打开按本服务预筛的规则清单	第 11.1 节至第 11.7 节
API 服务	「新建服务」填写上游端点 → 配置鉴权 → 发布并设置可用范围；成员在能力中心生成访问令牌	第 11.10 节
外部智能体	「接入外部智能体」→ 连通测试 → 注册后在「Agent 目录」中「批准」→ 配置上游凭据 → 发布并设置可用范围；按需「发布为 MCP 工具」	第 12.1 节至第 12.4 节
技能	「添加技能」→「分析来源」→ 核对风险等级与准入结论 →「确认安装」→ 发布并设置可用范围	第 12.5 节
插件	「导入」插件包或「导出为插件」；插件包不含凭据	第 12.6 节
外部注册中心	「新建源」接入 Nacos →「测试连接」→「立即同步」；同步进来的条目仍需单独分发	第 12.7 节

「发布」只让服务出现在能力中心目录中，不授予使用权限；成员能否使用由可用范围及其有效期决定。MCP 服务地址的出站访问受平台级「SSRF 出站防护」约束，该防护默认关闭；开启后指向内网的 MCP 地址一律被拒绝。接入不受信的服务前，由平台管理员评估网络位置后开启。接入服务后的工具发现与连接测试会在全部在线网关节点上执行，并在当前页面直接展示逐节点测试结果与失败原因。

7. 护栏与内容检测

护栏分两处配置：

1. 在「护栏」中维护检测依据：「敏感数据」（内置规则与自定义正则）、「词库」（内置黑名单、自定义黑名单、白名单）、「问答」（必答题库）和「内部系统白名单」；「威胁情报」页签接入本租户的恶意库（公开源或自定义 HTTP 批量源，每个租户最多 8 个）并维护例外名单。需要模型护栏许可。

2. 在「流量规则」中新建或编辑规则，选择「＋ 添加动作」→「内容检测」，选择检测方式（PII 敏感数据、词库、必答、风险类目、工具调用防护），设置命中后的处理与「组件不可用时」的「护栏检测出错」处理。图文联合检测与多轮拼接诱导防护在同一动作的「YAML 编辑」中配置（action.guard.image、action.guard.multiTurn），工具调用防护是「内容检测」的一种检测方式（原 MCP 服务详情页的独立护栏已并入路由策略）。检查出站地址是否命中恶意库时，选择「＋ 添加动作」→「恶意来源」，勾选「检查出站地址是否命中恶意库」，处置选「拦截」或「仅记录」；不开启即不检查。需要动态路由许可。

护栏页只定义能识别什么，没有流量规则引用时不会对调用生效；恶意库同样只在流量规则开启恶意来源检查后下发到网关，误拦的域名或 IP 在「威胁情报」页签的例外名单中放行，或在规则命中详情中点「加为例外」。平台管理员没有威胁情报入口；从早期版本升级后，需要重新接入恶意库并在流量规则中开启检查。人名、地址、机构名需要「敏感信息识别」模型；风险类目与提示注入检测需要「内容安全」模型；必答需要「问答语义匹配」模型，均在「全局设置」中指定。上线前先用「敏感数据试运行」与流量规则的「验证规则」核对，再用测试调用在请求详情中核对结论。拦截发生时的通知在「告警规则」中配置。详见《Zentrix Gateway 用户手册》第 13 章、第 14 章。

8. 流量规则、网关扩展与域名

「流量规则」需要动态路由许可，租户管理员或安全管理员可以进入；「全局限额」与「缓存」页签另需租户管理权限，由租户管理员维护。

1. 准备应命中与不应命中的请求样例。
2. 在「规则清单」中「新建规则」或「导入规则模板」，在画布中添加匹配条件与处置动作。配置了转发目标或「放行」「拒绝」的规则结束匹配，其余为增强规则，继续评估后续规则。
3. 选择「保存规则」，在「确认变更」中核对字段差异、关联对象和命中顺序后「确认并应用」。
4. 用「验证规则」核对命中链与最终目标，需要时勾选「模拟主档失败」查看后备顺序；模拟不调用上游，之后必须用真实调用验证。
5. 在「全局限额」中先选择调用类型，再「新增限额」，「限制类型」可选「请求次数」「并发限制」「熔断」「非法尝试保护」「上游调用超时」，请求次数与并发限制还需选择「计数维度」。
6. 在规则详情中选择「查看命中的请求」，或在「概览 → 路由策略命中」中核对实际命中。

熔断在失败或慢调用达到阈值时暂时拒绝请求；故障转移在转发目标中配置备用档位，二者不能互相替代。「网关扩展」中安装的 WASM 插件以停用状态安装，完成配置后再启用，并通过「执行统计」核对运行结果。字段、默认值和取值范围见《Zentrix Gateway 路由策略用户手册》与《Zentrix Gateway 用户手册》第 15 章、第 16 章。

9. 用量、观测与排障

入口	解决的问题
「概览 → 网关总览」	所选时间与数据范围内的费用、Token、成功率、耗时与调用流向；点击排行或归因下钻到网关请求
「概览 → 网关请求」	单次调用的结果、失败阶段、责任方、路由结论、命中记录、上游尝试和报文

入口	解决的问题
「概览 → 对话」	按会话查看多轮调用、耗时、费用与诊断
「概览 → 路由策略命中」	各条路由规则的命中次数、放行与拦截
「会话诊断」	自动识别的问题、原因、证据与修复建议，处理后第 7 天自动复查
「操作审计」	谁在何时修改了什么配置；需要全链路审计
「系统运行监控」「出口网络」	平台管理员查看平台各组成部分的状态与公网出口质量

排查顺序：确定时间与时区，在「网关请求」中按结果、责任方或请求 ID 找到目标调用，打开详情判断原因属于护栏、限流与配额、路由、调用身份还是上游，再进入对应配置处理；修正后用同一调用方重新发起请求核对。调用记录用于观察趋势与归属，结算以供应商或企业计费系统的账单为准。详见《Zentrix Gateway 用户手册》第 17 章至第 21 章。

10. 告警、导出与审计

1. 在「告警规则 → 通知渠道」中「新建渠道」（钉钉、飞书、Webhook、企业微信或邮件），选择「发测试消息」，在接收端收到后选择「确认收到」。只有已验证的渠道可以绑定告警规则；邮件渠道需要先完成「SMTP 配置」。测试结果与失败分类直接显示在当前配置卡片/对话框内，不会一闪而过；LDAP、OIDC、IM 渠道、镜像源、威胁情报恶意库以及网关出口代理的连通测试同样在页面内常驻展示测试结果（出口代理测试会逐节点展示各网关节点经该代理的连通耗时与状态）。
2. 在「告警规则」页签启用内置阈值规则并调整阈值，或「新建告警规则」→「网关策略事件：拦截、故障转移等」。费用通知使用「月度成本饱和度」规则并填写「月度预算」。
3. 告警事件在「概览 → 网关告警」中「确认」「静默」「标记已解决」或「指派」。
4. 在「报表导出」中「新建导出」，选择用量汇总或完整调用记录、频率和去向（收件箱、对象存储或 Webhook）。
5. 在「操作审计」中按操作人、动作或资源检索，导出审计证据 CSV 或 PDF，并在「归档与校验」中查看完整性校验结果。

调用内容与调用记录默认永久保留，需要调整时由技术支持协助配置；分析存储与对象存储的容量需要按调用量规划；操作审计记录在线保留两年。详见《Zentrix Gateway 用户手册》第 22 章。

11. 平台维护（平台管理员）

工作	入口	要点
升级	「系统维护 → 系统更新」	上传 `.zpkg` 升级包（可用「复制获取链接」下载），通过升级前检查后开始；开始前先完成一次备份；升级完成后在「部署配置」手动执行「立即验收」

工作	入口	要点
备份与恢复	「系统维护 → 备份与恢复」	手动「立即备份」并「下载」转存；恢复只适用于同一版本与部署形态；密钥服务的密钥目录 `/etc/zenova/secret-service/keys` 纳入主机级备份
运行状态	「运行与排障 → 系统运行监控」「出口网络」	控制面、数据面与依赖的状态；公网出口的延迟与丢包
运行参数	「网关与网络 → 网关」的「过载保护」页签；「系统维护 → 部署配置」的「资源容量」	处理器与内存的高、低水位；控制面数据库最大连接数（保存后控制面重启）
部署配置	「系统维护 → 部署配置」	访问地址、对象存储、运维通知渠道、链路导出、入口端口与绑定网卡、数据依赖、最近一次验收
网关与入口	「网关与网络 → 网关」	「概览」「入口与证书」「节点」「过载保护」「出口判定」「出口代理」；公开访问地址与入口证书；在「节点」中打开节点详情，点「查看节点日志」查看平台运行日志
租户	「租户 → 租户管理」	创建租户，结算货币创建后不能更改；设置每月 Token 上限与存储上限

删除租户、重置用户密码、替换或删除网关证书与域名、修改保存后需重启的运行参数（过载保护、资源容量）、启用旧版 MCP 身份以及退役网关节点等高风险操作，均通过统一的二次确认对话框执行（对话框写明操作动作与目标资源名称；删除租户与退役网关节点还需输入资源名称确认）。

控制台不可用时，在安装主机上执行 `sudo zenova status` 查看各服务状态。变更前记录版本与配置，明确维护窗口与回退方案；变更后验证登录、模型调用、权限拒绝、告警与备份状态。详见《Zentrix Gateway 用户手册》第 23 章、第 24.7 节，部署形态见《Zentrix Gateway 部署与服务选型指南》。

12. 首次上线验收

对象	正常路径	拒绝或失败路径
组织与角色	各角色账号只看到职责范围内的菜单	无权账号看不到菜单，直接访问地址被拒绝
模型服务	试点成员获取 API 密钥并调用成功	可用范围外、Key 已撤销和额度耗尽的调用被拒绝
MCP 与外部能力	授权成员获取连接配置并完成只读调用	未分发成员无法签发 Token，受限工具被拦截

对象	正常路径	拒绝或失败路径
流量规则	「验证规则」与真实调用的路由结论一致	限额、熔断与反例按预期处理；规则清单没有被数据面拒绝的提示
护栏	正常内容通过	合成敏感内容按设定脱敏、记录或拦截
观测与审计	「网关请求」「路由策略命中」与「操作审计」中可查到对应记录	失败调用能定位到责任方与原因
运维	通知渠道测试成功，备份已转存	恢复已在隔离环境演练

验收记录包含版本、环境、请求 ID、预期与实际结果以及未完成项。分阶段上线的步骤与检查清单见《Zentrix Gateway 用户手册》第 25 章、第 26 章。