FEATURE SPECIFICATION

功能说明手册

统一智能运维监控平台 · 三个版本 · 完整功能矩阵

本手册面向技术选型、POC 验证与实施交付，逐项说明各功能域的能力范围、版本归属与使用限制， 并给出部署规格、集成方式与授权机制。适用于图班.浑仪 1.9.4 及 V3.0 版本体系。

文档版本	适用软件	适用版本体系	更新日期
V3.0	1.9.4	社区 / 企业 / OEM	2026-11

0 关于本手册

本手册是产品能力的权威描述，用于技术评估与验收比对。功能是否存在、在哪个版本可用、受什么限制，以本手册为准；宣传材料与本手册冲突时，以本手册为准。

0.1 阅读对象

技术选型
看第 3–5 章：功能域能力、版本矩阵、配额限制。

实施交付
看第 7–8 章：部署规格、集成方式。

运维管理
看第 9–10 章：安全合规、授权与升级。

0.2 标记约定

标记	含义
✓	该版本包含此能力，无额外限制
◐	限量 / 限时可用：以试用额度或配额形式提供，超出后需升级
—	该版本不含此能力：界面保留入口并显示升级引导，API 返回 `FEATURE_NOT_LICENSED`

0.3 目录

1 产品定位与版本体系 03
2 系统架构 04
3 功能域总览 05
4 功能详述 06
5 三版功能矩阵 12
6 配额与限制 14
7 部署规格 15
8 集成能力 16
9 安全与合规 17
10 授权与升级 18
11 附录 19
– 联系方式 19

IMPORTANT

单一代码库，差异全在授权。三个版本是同一套安装包、同一套镜像，功能差异完全由授权文件决定。 从社区版升级到企业版只需激活新的授权，**无需重装、无需迁移数据、无需停机**；降级同样平滑，已有数据与配置完整保留。

1 产品定位与版本体系

1.1 产品定位

图班.浑仪是面向私有化环境的统一智能运维监控平台，覆盖云平台、服务器主机、网络设备三类资产域，提供从资产纳管、指标采集、告警治理到 AI 根因分析、处置编排与合规报告的完整闭环。默认单机内网部署，数据不出网。

1.2 三个版本的划分逻辑

版本不按"功能多少"简单切分，而按使用形态划分——这是判断某个能力归属哪一档的标准：

版本	回答的问题	分界线	典型客户
社区版	一个人能不能把这套环境管起来	单人自用	个人运维、homelab、30 台以内小环境
企业版	一个团队能不能靠它值班、合规、对外集成	组织协作 + 规模	中大型企业 IT 部门，100 台起
OEM 版	别人能不能贴自己的牌去交付、自己发证	渠道交付	集成商、信创项目、大客户定制

1.3 五道能力闸门

社区版开放约七成的功能面，覆盖单人完整监控闭环；以下五类能力构成版本闸门：

闸门	包含能力	设置理由
A 团队协作	邮件 / 企微应用通知、值班排班、静默窗口、依赖抑制（短信通道规划中）	告警送达是监控产品的最低可用标准；该能力随用的人数变多而价值上升
B 规模	资产档位、采集分片多实例	规模带来真实的采集与存储成本，按量收费
C 合规治理	基线合规（CIS / 等保）、多实例高可用（规划中）、PostgreSQL 指标层	受监管客户的刚需，与规模无关的治理投入
D 平台集成	开放 API / 事件订阅、云平台纳管、K8s、LLDP 自动拓扑、RAG 知识库	客户进入"平台化"阶段时预算才成立
E 渠道交付	编译期换牌、信创构建、子授权、源码级定制、专属 SLA	仅对交付方有价值，最终用户无需此能力

1.4 升级与降级

- **升级**：激活新的授权文件，功能与配额即时生效，无需重启服务、无需重装、数据不动。
- **降级**：授权到期或机器指纹不匹配时，自动回退社区功能集。**不锁数据、不停采集、不删配置**——超出配额的部分仅禁止新增。
- **试用**：企业版可签发 30 天全功能授权，不绑定机器指纹，便于先装后买；到期自动降级。

DESIGN NOTE

授权失效不以"客户生产环境瘫痪"的方式处理。过期、指纹不匹配、授权文本损坏，一律降级为社区版继续运行， 仅收敛付费功能。这是产品的明确设计取向，不是配置选项。

2 系统架构

2.1 组件构成

平台以两个容器交付：`api`（Web 服务与 API）与 `collector`（采集执行器）。两者共享同一数据卷，通过本地数据库协作，无外部中间件依赖。



2.2 关键设计

设计	说明
零外部依赖	不依赖 PostgreSQL、Redis、Kafka 等外部组件即可运行；数据落在本地单文件数据库，备份即拷文件。规模上来后可启用 PostgreSQL / TimescaleDB 指标层（企业版）。
采集与展示分离	采集在独立容器内执行，采集器异常不阻塞 Web 与告警查询；采集器自身有心跳与水位自监控。
三级降采样	原始点 → 1 分钟聚合 → 1 小时聚合。查询按时间跨度自动选层，长周期查询不拖慢控制台，同时保留期不因降采样而缩短。
单文件数据	全部业务数据在一个数据库文件内，一致性由事务保证；跨容器访问通过文件锁协调，升级时先停容器再迁移。
凭据加密	设备口令、SSH / BMC 凭据、通知渠道密钥、大模型 Key 均加密落库，密钥可由环境变量或独立密钥文件提供。

2.3 端口与路径

项	默认值
Web / API 端口	`18080`（HTTP，可映射到自定义宿主端口）
健康检查	`GET /health` 返回服务与数据库状态
数据目录	`/data/argus/data`（数据库文件所在，需持久化）
密钥目录	`/data/argus/secrets`（启用密钥外置后，仅服务账号可读）
部署目录	`/data/argus`（编排文件、环境变量、安装脚本）

3 功能域总览

平台能力按十四个功能域组织，下表为索引；逐项说明见第 4 章。

#	功能域	核心能力	版本归属
1	运行总览与告警大屏	全局健康度、告警水位、资产分布、大屏指挥视图	三版均含
2	资产纳管	主机 / 网络 / 云资源 CRUD、批量导入、连接测试、机架 U 位、BMC、应用服务拨测	三版均含（配额不同）
3	采集	ICMP / SNMP / gNMI / GPU(DCGM) / Exporter / Agent 主动采集，Trap / Syslog 被动接入	三版均含（分片为企业版）
4	指标存储与查询	三级降采样、时序查询、仪表盘、资源详情	三版均含（PG 指标层为企业版）
5	告警	规则引擎、状态机、收敛抑制聚合、风暴抑制、值班、通知渠道	核心三版均含；邮件 / 值班为企业版（短信规划中）
6	拓扑	轻量拓扑推断、上下游关系、根因候选	轻量三版均含；自动发现为企业版
7	AI 分析与助手	规则引擎兜底、多厂商大模型、证据链根因、运维助手	规则兜底三版；大模型按额度
8	知识库	故障知识沉淀、检索、RAG 知识复利	基础检索三版；RAG 为企业版
9	运维报告	日 / 周 / 月报、优化建议、HTML / MD / JSON / PDF 导出	三版均含
10	处置编排与审批	声明式剧本、演练、审批流、执行留痕、回滚	三版均含
11	基线合规	CIS / 等保技术要求核查、定向检查、差距清单	企业版 / OEM
12	运维工具	诊断、维护窗口、备份还原、变更记录	三版均含（备份策略相同）
13	用户权限与组织	三角色 RBAC、部门组织、审计日志	三版均含（用户配额不同）
14	系统与授权	品牌设置、授权管理、通知配置、系统参数	品牌定制为企业版；换牌为 OEM

NOTE

"三版均含"指该能力在社区版即可使用，不受授权限制。带配额差异的项在第 6 章列明具体数值。

4.1 运行总览与告警大屏

登录后的默认视图，用于值班场景下的第一眼判断：整体健康度、当前活跃告警数、按严重级别分布、资产在线率、采集器状态。大屏视图面向指挥中心与电视墙，**约 15 秒刷新水位**，在 1000+ 资产规模下实测 40 秒内仅产生 2 次查询请求，不会给数据库造成压力。

- 健康度按资产与告警加权计算，可下钻到具体资产
- 告警水位、趋势、Top N 资产在单屏内可见
- 大屏适配 16:9 与 21:9，支持无操作自动轮播

4.2 资产纳管

统一管理主机、网络设备与云上资源。资产是平台的核心对象，采集、告警、拓扑、合规核查全部围绕资产展开。

能力	说明
资产模型	名称、类型、厂商、型号、OS、管理地址、站点、机房、机架与 U 位、责任人、部门、环境、标签
机房 / 机架 / U 位	新建或编辑资产时先选机房，机架编号输入框自动联想该机房内已有编号（机架编号仅在机房内唯一，换机房候选随之收窄）；U 位自由填写。资产详情页展示「机架 · U 位」，资产列表可按机房、机架筛选定位。无机柜正视图：机架与 U 位以字段记录与筛选为准，不做机柜可视化绘图；应用类型资产不涉及这三项
批量导入	Excel 模板导入，导入前校验必填项与地址合法性，失败行单独回执
连接测试	保存前验证连通性与凭据有效性，避免"录进去才发现不通"
凭据管理	SSH、SNMP、BMC 口令加密存储；凭据可复用于多台资产，轮换时一处更新
BMC 带外	纳管服务器带外管理口，硬件健康、电源、风扇、温度等状态独立于 OS 采集
维护窗口	为资产设定维护期，期间抑制相关告警并记录原因
组织归属	按部门 / 站点划分，支持按归属过滤与分权查看

4.3 采集

采集在独立容器内执行，支持主动轮询与被动接入两条路径，采集结果统一进入指标层与事件流。

方式	说明
ICMP	连通性与时延探测，作为可用性基础指标
SNMP	网络设备与支持 SNMP 的主机指标轮询，按 profiles 定义采集项
gNMI	网络设备流式遥测（gRPC）；内置主流厂商私有模型模板，自动补 CPU / 内存 / 温度路径与单位换算
Exporter	拉取 Prometheus 格式的指标端点，适配已有 exporter 生态
GPU / DCGM	经 dcgm-exporter 采集 GPU 卡级硬件状态与利用率，配套 5 条按卡判断的内置硬件告警（致命 XID / ECC / 行重映射 / 热降频）
Agent	在托管主机上运行轻量代理，采集主机内部指标
Simulator	内置模拟数据源，用于演示、培训与剧本演练，不接触生产设备
SNMP Trap	被动接收设备主动上报的事件，与轮询指标共用时间线
Syslog	接收网络设备与主机日志，作为事件源参与告警与根因关联

GPU 采集器支持**纳管时经 SSH 自动部署**：建档勾选后由后台任务安装 dcgm-exporter（含 `--restart` 与自定义字段清单），失败按「驱动缺失 / 缺 toolkit / 镜像拉取失败」分级回执；也可复制等价部署命令手工执行。采集器具备自监控：任务成功率、最近采集时间、单次耗时、失败原因均可在界面查看；单台资产采集异常不会影响其他资产。

4.4 指标存储与查询

- **三级降采样**：原始点保留短期精度，1 分钟与 1 小时聚合层支撑长周期查询。查询按时间跨度自动选层，用户无需关心。
- **保留期不缩水**：降采样是为了查询性能，不用于压缩历史保留。社区版与企业版的保留策略一致。
- **指标层可选**：默认使用内置存储；企业版可切换至 PostgreSQL / TimescaleDB，支撑更大规模与更复杂查询。
- **仪表盘**：按资产或指标维度组合图表，支持自定义时间范围与对比。

4.5 告警

告警引擎由规则引擎与状态机组成，目标是"该报的报、不该报的不报"，而不是把阈值命中全部推给人。

能力	说明	版本
规则引擎	按指标、阈值、持续时长、组合条件定义规则；支持针对资产组与标签批量套用	三版
状态机	告警有明确的触发 / 持续 / 恢复 / 关闭状态，避免抖动导致反复通知	三版
收敛与聚合	同源告警合并、按时间窗聚合，把"一次故障一次通知"落到位	三版
抑制	上游故障抑制下游衍生告警；维护窗口内抑制；依赖关系抑制	三版（依赖抑制为企业版）
风暴抑制	单位时间内告警量超阈值时自动限流并标记，防止通知通道被打爆	三版
规则试算	规则上线前用历史数据试跑，看命中量是否合理，避免"上线即刷屏"	三版
值班排班	按班次指定值班人，告警按当前值班人分派与升级	企业版
静默窗口	计划性静默，带审批与到期自动恢复，防止"静默了忘记打开"	企业版

通知渠道

渠道	说明	版本
站内通知	实时推送至控制台，带未读计数	三版
机器人 Webhook	企业微信 / 钉钉 / 飞书群机器人，配置简单，适合个人与小团队	三版
邮件 SMTP	标准 SMTP 发送，支持 TLS；有投递记录便于排查	企业版
企微应用消息	通过企业微信应用发送，可指定接收人与部门	企业版
个人微信	Server酱 / PushPlus 公众号中转，按人推送；HTTP 200 不代表送达，按业务码判定，详见 4.17	企业版
短信 / 电话	面向 critical 级告警的强触达通道	企业版（规划中，按项目启动）

4.6 拓扑

- **轻量拓扑**（三版）：基于同机房、同网段、资产归属等已知关系推断连接，生成可视化拓扑，无需设备侧配合。
- **自动发现**（企业版）：通过 LLDP / CDP 邻居关系自动构建物理拓扑，并对根因候选进行打分排序。
- 拓扑参与告警抑制：上游设备故障时，下游设备的衍生告警自动收敛，减少噪音。

4.7 AI 分析与运维助手

采用"规则兜底 + 大模型归因"的双层结构。即使不接任何外部模型，平台仍具备基于规则的故障判断能力——这是社区版也能用的部分。

能力	说明	版本
规则引擎兜底	内置常见故障模式的判断逻辑，离线可用，不产生外部调用	三版
大模型归因	汇总指标拐点、同期事件、变更记录、拓扑关系，产出根因排序与处置建议	按额度
证据链	每条结论附带可展开的原始数据引用，供人工复核	按额度
多厂商接入	支持主流公有云与私有化部署模型，客户自带 Key，token 成本归客户	按额度
运维助手	自然语言问答，查资产、查告警、解释指标含义	按额度

AI 调用按自然月计数，配额用尽后自动回落至规则引擎兜底，不影响监控与告警主流程。

4.8 知识库

- **知识沉淀**（三版）：把处置过程、根因结论、复盘要点存为条目，支持标签与全文检索。
- **RAG 知识复利**（企业版）：知识条目参与 AI 检索增强，同类故障再次发生时可直接命中历史结论，模型回答附带知识出处。

4.9 运维报告

按日 / 周 / 月自动生成运维报告，内容包括资产健康概览、告警统计与 Top 问题、趋势对比、优化建议。支持 HTML、Markdown、JSON 与 PDF 四种导出格式，PDF 可直接用于汇报与归档。

- **PDF 开箱即用**：镜像内置无头 Chromium 与中文字体，离线环境导出 PDF 无需另装浏览器，中文正常显示。
- 报告模板可调整章节与时间范围，适配周会与月度汇报两种节奏
- JSON 导出便于对接自有报表系统或做二次分析
- 报告生成不占用在线查询资源，历史报告可回溯查看

4.10 处置编排与审批

把"发现问题之后怎么办"固化下来：处置动作写成声明式剧本，经审批后由平台代为执行，全过程留痕可回滚。

能力	说明
声明式剧本	定义前置条件、执行步骤、成功判据与回滚动作；剧本是配置而非临时脚本，可版本化
编排执行	按步骤顺序执行，任一步失败可按策略中止并触发回滚
审批流	执行前提交审批，高危动作强制二次确认；审批人与执行人可分离
演练模式	在模拟数据源上完整跑一遍剧本，确认逻辑正确再用于生产
执行留痕	每次执行记录触发人、审批人、步骤耗时与输出，写入审计日志

4.11 基线合规

对主机做配置类安全基线的自动核查，输出可交付的差距清单。**支持按主机定向检查**——只勾选需要核查的那几台，只跑这几台，结果页与报告中明确标注检查范围。

能力	说明	版本
基线条目	覆盖 CIS 与等保技术要求中的配置类条目，如账号策略、口令强度、SSH 配置、审计开启、权限位等	**企业版**
定向检查	选择指定主机范围执行，支持搜索、多选、全选；空选即全量	**企业版**
结果追溯	每次检查记录范围、时间、执行人；历史记录中的范围信息不可篡改	**企业版**
差距清单	逐项给出通过 / 不通过、风险说明与修复建议，可导出用于整改	**企业版**

4.12 运维工具

诊断工具

对资产做连通性、端口、协议、凭据的分层诊断，定位"采不到"的具体原因。

维护窗口

计划性维护登记，期间抑制告警并记录原因，事后可核对。

备份还原

每日自动备份，支持手动触发与还原；单文件数据使备份与迁移成本极低。

4.13 用户权限与组织

能力	说明
三角色 RBAC	管理员 / 运维 / 只读三类角色，权限边界清晰；菜单与接口双重校验，前端隐藏不作为唯一防线
用户配额	按授权限制可创建的用户数，超额禁止新增，不移除已有用户
组织与部门	维护组织结构，资产可按部门归属，支持按组织范围过滤数据
认证	账号口令认证，口令单向哈希存储（至少 10 位且含字母与数字）；管理员新建账号或重置口令后，本人首次登录必须修改口令才能使用其它功能，未改密时除「查看本人信息 / 改密 / 退出」外的接口一律拒绝
审计日志	登录、配置变更、授权变更、剧本执行、静默操作全部留痕，含操作人、时间、对象与结果

4.14 系统设置与授权

能力	说明	版本
授权管理	查看当前版本、授权主体、有效期、配额与已用量；激活新授权即时生效	
品牌设置	系统名称、主题色、深色模式	三版
LOGO 与标识定制	替换 LOGO、隐藏厂商标识	企业版
完全换牌	编译期品牌注入，二进制层面不含厂商标识	OEM
通知配置	SMTP、机器人通道参数配置与测试发送（短信通道规划中）	
系统参数	采集间隔、保留策略、告警默认值等运行参数	

约束

版本与配额只能来自授权文件，界面不提供手动切换版本的入口。 这是为了防止"改一个下拉框就完成升级"。任何版本判断都以后端验签结果为准，前端展示仅为渲染需要。

4.15 云平台与 K8s（企业版 / OEM）

能力	说明	版本
vSphere 纳管	接入 vCenter，自动纳管 ESXi 主机与云主机清单；跟踪云主机电源状态（关机自动标维护、下线标记不删除）	企业版
vSphere 指标	CPU / 内存使用率与 20 秒级磁盘 / 网络 IO（quickStats）；vCenter 与 ESXi / 云主机建立从属拓扑	企业版
公有云纳管	阿里云 ECS 按地域翻页拉取全量实例（名称 / 规格 / IP / 可用区 / 标签 / 状态），账号内按实例 ID 去重；只读 RAM 子账号授权	企业版
自动发现	按纳管范围自动生成资产，新增虚拟机自动纳入监控；资产配额防超建	企业版
K8s 集群	kubeconfig 或 API Server + ServiceAccount Token 双接入；自动纳管节点，Pod 重启、Deployment 副本、Warning 事件（同类 5 分钟去重）作为告警源	企业版
统一视图	云上资源与物理资产在同一资产视图与拓扑中呈现	企业版

注：**vSphere 纳管、阿里云 ECS 与 K8s 集群监控已交付**，随企业版 / OEM 提供；OpenStack / ZStack 等其余云平台按项目定制排期。

4.16 应用服务监控

面向部署在机房里的业务系统（Web 服务、中间件、自研程序）：不进业务代码，从外部验证"服务还活着、响应是否正常"。应用资产挂在一台**承载主机**上，由该主机的 Agent 周期执行检查；Agent 离线期间检查暂停，恢复后自动继续。

检查项	说明
HTTP 拨测	期望状态码（默认 200）、响应关键词、超时时间、Bearer 令牌头、跳过证书校验（自签证书的内网服务可用）
进程检查	按进程名关键字匹配，单应用最多 32 个进程名；进程消失即判异常
数量上限	单应用最多 16 个检查项，可逐项启用 / 停用
快捷配置	承载主机的 Agent 已上报监听端口时，可从监听端口候选一键带出 URL 与期望状态码

指标与呈现

指标	含义
app.http.up	本次拨测是否成功（1 / 0）
app.http.status_code	HTTP 响应状态码
app.http.response_ms	响应耗时（毫秒）
app.http.cert_days_left	HTTPS 证书剩余天数，到期前可配告警提前换证

- 界面路径：**资产与采集** → **应用服务**；新建类型为"应用"的资产 → 绑定承载主机 → 配置检查项。
- 应用详情页提供可用率曲线、检查项明细（每次检查的状态码 / 耗时 / 命中关键词）与承载主机跳转。
- 告警复用平台统一告警引擎：检查失败按规则产生告警，走 4.5 节的通知渠道触达。
- 三个版本均含，占用资产配额。

4.17 个人微信通知（Server酱 / PushPlus）

面向"值班人不在企业微信群里"的场景：告警经公众号中转，直接推送到个人微信。

渠道	说明
Server酱	SendKey 从 sct.ftqq.com 获取；兼容 Server酱³ 的 `sctp{uid}t{key}` Key 格式；支持自定义 baseUrl（私有化中转）
PushPlus	token 从 pushplus.plus 获取；支持群组 topic（一次配置，组内多人接收）
送达判定	中转接口返回 HTTP 200 不代表送达，平台按业务码判定成败并留投递记录，便于排查"发了但没收到"

- 界面路径：**告警** → **通知渠道** → **新建渠道**，渠道类型选"个人微信"，填 SendKey / token 与接收人。
- 属企业版能力（与企业微信应用消息同一功能位）；密钥加密落库，界面回显脱敏。

5 三版功能矩阵

✓ 包含 ◐ 限量 / 限时 — 不含

功能点	社区版	企业版	OEM 版
运行总览 / 告警大屏	✓	✓	✓
资产纳管（主机 / 网络 / 云）	✓	✓	✓
资产配额	**30**	100 / 500 / 2000 / 不限	不限
批量导入 / 连接测试 / 机架 U 位	✓	✓	✓
BMC 带外管理	✓	✓	✓
ICMP / SNMP / Exporter / Agent	✓	✓	✓
gNMI 流式遥测 / 厂商模板	✓	✓	✓
GPU / DCGM 监控与纳管自动部署	✓	✓	✓
SNMP Trap / Syslog 被动接入	✓	✓	✓
采集分片多实例（大规模）	—	✓	✓
三级降采样 / 仪表盘 / 资源详情	✓	✓	✓
PostgreSQL / TimescaleDB 指标层	—	✓	✓
告警规则引擎 / 状态机	✓	✓	✓
收敛 / 聚合 / 抑制 / 风暴抑制	✓	✓	✓
规则试算	✓	✓	✓
站内通知 + 机器人 Webhook	✓	✓	✓
邮件 SMTP	—	✓	✓
企业微信应用 / 个人微信（Server酱 / PushPlus）	—	✓	✓
短信 / 电话通知	—	**◐ 规划中**	**◐ 规划中**
值班排班 / 静默窗口	—	✓	✓
拓扑依赖抑制	—	✓	✓

5 三版功能矩阵（续）

✓ 包含　◐ 限量 / 限时　— 不含

功能点	社区版	企业版	OEM 版
轻量拓扑推断	✓	✓	✓
LLDP / CDP 自动发现	—	✓	✓
规则引擎兜底分析	✓	✓	✓
大模型分析（自带 Key）	◐ 100 次 / 月	不限	不限
RAG 知识库	—	✓	✓
运维报告（HTML / MD / JSON / PDF）	✓	✓	✓
应用服务监控（HTTP 拨测 / 进程检查）	✓	✓	✓
剧本编排 / 演练 / 审批	✓	✓	✓
基线合规（CIS / 等保）	—	✓	✓
诊断 / 维护窗口 / 备份还原	✓	✓	✓
用户配额	5	不限	不限
三角色 RBAC / 审计日志	✓	✓	✓
品牌设置（名称 / 主题）	✓	✓	✓
LOGO 定制 / 隐藏厂商标识	—	✓	✓
编译期完全换牌	—	—	◐ 渠道合作后提供
开放 API / 事件订阅	—	✓	✓
多实例 / 高可用	—	◐ 规划中	◐ 规划中
vSphere 纳管 / 阿里云 ECS 纳管	—	✓ 已交付	✓ 已交付
K8s 集群监控	—	✓ 已交付	✓ 已交付
其余云平台（OpenStack / ZStack 等）	—	◐ 按项目定制	◐ 按项目定制
信创构建（麒麟 / 统信、达梦 / 金仓）	—	◐ 按单加购	✓
子授权（自行签发最终客户）	—	—	◐ 渠道合作后提供
源码级定制 / 驻场 / 专属 SLA	—	—	✓

6 配额与限制

6.1 配额定义

维度	社区版	企业版	说明
资产数	**30**	100 / 500 / 2000 / 不限	按授权档位签发，可在授权中单独调整
用户数	**5**	不限	社区版按 5 个账号计
AI 月调用	**100 次**	不限（Key 自备）	自然月重置；Key 由客户自备，不售卖 token，次数仅防滥用
数据保留	与付费版一致		保留策略不因版本降级而缩短

6.2 超额处理：拒新增、保存量

生产环境超出配额时直接拒绝新增，存量行为完全不受影响：

行为	结果
新增资产 / 用户	拒绝并返回 `QUOTA_EXCEEDED`，界面给出已用 / 上限
已有资产采集	**继续正常采集**，不因超额停采
已有数据	**不删除**，历史指标、告警、报告全部保留
界面提示	顶部常驻配额提示条，接近上限时变为警示色，点击进入授权页
审计记录	配额告警按维度限频写入审计（同一维度 10 分钟内一条），避免刷屏

为什么是拒新增

超配额只挡新增、不动存量：已有资产的采集、历史数据、报告全部保留，客户感受到的是"该扩容了"， 而不是"这破软件又崩了"。开发与演示环境（`LICENSE_ENV` 未设为 production）只写审计告警、不做拦截， 避免本地老库把使用者锁在外面。

6.3 其他限制

项	限制
授权有效期	社区版永久；企业版按订阅期；试用 30 天
机器绑定	正式授权可绑定机器指纹（16 位短哈希），换机需重新签发；试用授权不绑定
单实例部署	社区版与企业版标准档为单实例；多实例 / 高可用为企业版能力（规划中，按项目启动）
并发查询	默认按单机规格设计，超大规模建议启用 PostgreSQL 指标层并调高资源配置

7 部署规格

7.1 环境要求

项	要求
操作系统	Ubuntu 20.04 / 22.04 / 24.04；信创环境走麒麟 OS / 统信 UOS 适配构建（OEM / 企业版加购）
容器运行时	无需自备：交付包内置 Docker 静态引擎与 compose 插件，未安装时由安装脚本解包部署并注册 systemd 服务，全程不出网；客户已装 Docker 时自动跳过
架构	x86_64；ARM64 / LoongArch 随信创构建提供
网络	默认内网运行，除安装与升级外无需出网；被纳管设备需开放对应采集协议

7.2 资源规格建议

规模	CPU	内存	磁盘	说明
≤ 100 资产	2 核	2 GB	20 GB	最低配置，单机即可
100 – 500	4 核	8 GB	100 GB	建议独立数据盘
500 – 2000	8 核	16 GB	500 GB	启用 PostgreSQL 指标层，采集分片
> 2000	多实例部署评估			需结合采集频率与保留期做容量规划

磁盘估算主要取决于指标保留期与采集粒度；建议保留期不低于 30 天，按实际增长率调整。

7.3 安装流程

- **在线安装**：执行一条 `curl ... | bash` 命令，脚本自动完成环境检查、包下载、SHA256 校验、镜像载入、密钥生成、数据库初始化、创建管理员、启动与健康检查。下载镜像包时**实时显示进度**（百分比 / 速率 / 剩余时间），中途断线不留半截文件，重跑即可续装。
- **离线安装**：交付一个整包 `.tar.gz`，内含镜像、编排文件、安装脚本、**Docker 静态引擎（`docker-<版本>.tgz`）与 compose 插件（`docker-compose`）**、密钥迁移脚本、部署手册与校验清单。解压后本地执行安装脚本，流程与在线一致。
- **断网可装**：客户机器未装 Docker 时，安装脚本把包内自带的静态引擎解包到 `/usr/local/bin` 并注册 systemd 服务，全程不出网；机器已装 Docker 时自动跳过，不覆盖现有环境。
- **校验**：安装前按 `SHA256SUMS` 逐文件校验；缺失文件自动补齐后重校，校验不通过即中止，不进入部署阶段。
- **初始化**：自动生成数据库与服务密钥，创建初始管理员账号，终端打印访问地址与初始口令。

7.4 升级与回滚

操作	说明
原地升级	指定新版本重跑安装脚本。升级前自动停容器、备份数据库、执行迁移，再启动新版本。保留配置、密钥、账号与全部数据。
停机时间	迁移期间需独占数据库，容器停止时间通常数十秒，与数据量相关。
回滚	切回上一版本镜像即可；数据库不自动回滚，如需回退数据结构，使用升级前的备份还原。
密钥轮换	提供密钥外置与轮换脚本：将服务密钥从环境变量迁至独立文件，并用新密钥重新加密库内凭据，执行后环境变量与容器检查中不再出现明文。

8 集成能力

8.1 对外接口

方式	说明	版本
开放 API	REST 接口（`/api/v1/open/*`）：whoami 自检、资产清单、指标曲线、告警读与认领；令牌 + scope（resources:read / alerts:read / alerts:ack / metrics:read），权限 = scope ∩ 创建人当前角色权限；单令牌 60 次/分钟限流	**企业版**
事件订阅	告警产生 / 恢复、资产新增三类事件推送到回调地址；HMAC-SHA256 签名（`X-Argus-Signature`），失败按 1 分钟–6 小时五档退避重试，6 次失败入死信、可控制台重放；投递记录可查	**企业版**
报告导出	HTML / Markdown / JSON / PDF，可定时推送至指定位置	**三版**
机器人 Webhook	企微 / 钉钉 / 飞书群机器人，配置即用	**三版**

开放 API 与事件订阅**已交付**：接入指南、验签示例与错误码见随包交付的《开放接口对接文档》（`docs/open-api.md`），机器可读描述见 OpenAPI 3.1 文件（`docs/openapi/argus-open-api-3.1.yaml`）。

四步接入（控制台 **系统** → **开放接口**，需 `settings:write` 权限）：

- ① **签发令牌**：在「API 令牌」页签新建，勾选 scope（如工单系统勾 `alerts:read` + `alerts:ack`）；密钥明文只显示一次。
- ② **复制示例自检**：签发后页面示例**自动把占位符换成刚签发的真令牌**，点「复制」即可直接 `curl` 调 whoami / 资产 / 告警接口验证连通。
- ③ **配事件订阅**：在「事件订阅」页签填回调地址，平台生成 HMAC 密钥（同样只显示一次，可随时轮换）；告警产生 / 恢复、资产新增即推送到客户系统。
- ④ **看投递记录**：投递失败自动按五档退避重试，6 次失败入死信；在投递记录页签查看每次推送明细并可一键重放。

8.2 被集成方向（数据接入）

来源	接入方式
Prometheus 生态	以 Exporter 拉取方式接入已有指标端点
网络设备	SNMP 轮询 + Trap 被动接收
主机	Agent 代理采集，或通过 SSH 做配置与基线核查
日志	Syslog 接收，作为事件源参与告警与根因关联
云平台 / K8s	纳管式接入，自动同步资源清单（企业版；vSphere / 阿里云 ECS / K8s 已交付，其余按项目定制）

8.3 典型集成场景

场景	做法
与 CMDB 对接	已交付：开放 API 拉取资产与告警，CMDB 为权威源时由平台定时拉取增删
与工单系统联动	已交付：事件订阅推送告警到工单系统自动建单，处置完成后调认领接口回写状态
与统一门户集成	大屏或总览以嵌入方式挂到门户，走单点登录或免登只读视图
与值班系统联动	通过机器人通道把告警推到现有值班群，无需替换原有流程

接口约定

接口版本随产品版本演进，向后兼容优先；破坏性变更会在大版本中说明并提供迁移期。 鉴权使用 `X-API-Key` 请求头（也接受 Bearer 方式），令牌权限与平台内角色权限求交，不会绕过 RBAC。

9 安全与合规

9.1 运行安全

项	措施
运行身份	容器以非 root 的最小权限用户启动，不以特权模式运行
配置隔离	环境变量文件权限 600；启用密钥外置后，密钥目录仅服务账号可读
数据落盘	数据库文件在数据目录内，设备口令、渠道密钥、模型 Key 加密存储
网络暴露	默认仅监听一个 Web 端口，建议只对管理网段开放，不对公网暴露
镜像内容	业务代码以编译产物交付，镜像内不含源码与调试符号
完整性	安装包按清单逐文件校验；交付物内置完整性校验与批次标识

9.2 访问控制

- 三角色 RBAC：管理员、运维、只读；菜单与接口双重校验，前端隐藏不作为唯一防线。
- 口令单向哈希存储，不保存明文；口令强度要求为至少 10 位并同时包含字母与数字。
- 首次登录强制改密：管理员创建的账号、被管理员重置过口令的账号，登录后必须改成本人口令；该限制在服务端强制（其它接口返回 403），前端页面仅作引导。
- 会话与令牌由服务端校验，越权请求返回明确的权限错误。
- 审计日志覆盖登录、配置变更、授权变更、剧本执行与静默操作，记录操作人与时间。

9.3 合规支持

要求	平台支持方式	版本
等保 2.0 技术要求	基线核查覆盖配置类条目，输出差距清单与修复建议	**企业版**
CIS Benchmark	按 CIS 条目逐项核查，结果可导出用于测评留档	**企业版**
操作审计留痕	关键操作入审计日志，可查询与导出	**三版**
数据本地化	数据完全落在客户环境，默认不向外部回传	**三版**
信创环境	麒麟 / 统信 OS、达梦 / 人大金仓数据库适配构建	**可选 / OEM**

9.4 加固交付说明

交付物采用编译交付与完整性校验组合：业务代码不以可读源码形式进入镜像，交付物内置批次标识与启动校验；服务密钥支持移出环境变量、改为最小权限文件托管，并支持一次密钥轮换，使容器配置检查中不再出现明文密钥。

具体加固措施与部署建议，以随包交付的加固说明文档为准。

10 授权与升级

10.1 授权模型

版本与配额完全由授权文件决定，授权文件由厂商签发，包含版本、授权主体、有效期、机器指纹绑定信息与配额。平台在启动与运行期校验授权有效性，校验结果决定生效的版本。

要素	说明
签发	由厂商使用私钥签发，客户侧仅持有验证所需公钥，无法自行伪造或修改
绑定	可绑定机器指纹，授权只能在指定主机生效；试用授权不做绑定
有效期	支持固定期限与永久两种；到期前可在界面看到剩余天数
配额	资产、用户、AI 月调用三项随授权下发，可针对单个客户单独设定

10.2 授权状态与行为

状态	含义	平台行为
未激活	没有授权文件	以社区版运行，功能与配额按社区版
有效	校验通过且在有效期内	按授权版本与配额运行
已过期	超过有效期	降级社区版，数据与配置完整保留，界面提示续期
指纹不匹配	授权绑定其他主机	降级社区版，提示联系厂商重新签发
历史授权	旧版本手填的授权信息	仅展示，不参与校验；提示重新激活
文件损坏	授权文本被篡改或损坏	视同未激活，降级运行而非报错中断

10.3 换档流程

- 在「授权与升级」页复制本机机器指纹，提交给厂商。
- 厂商按目标版本与配额签发授权文件。
- 粘贴授权文本完成激活，**即时生效，无需重启或重装。**
- 降档同理：激活低一档授权即收敛功能与配额，存量数据不受影响。

10.4 试用

企业版可签发 30 天全功能试用授权，功能与企业版一致，不绑定机器指纹，便于客户在自有环境完整验证。到期后自动降级为社区功能集，试用期间产生的数据、配置、账号全部保留。

> 采购建议
>
> 若不确定档位，先用社区版跑两周：看实际纳管资产数、值班人数、是否需要邮件触达、是否要交合规报告。 这四项决定了起点档位，避免买超或买少。

11 附录

11.1 术语表

术语	含义
资产	被纳管的对象，可为主机、网络设备、云上资源等；采集、告警、合规均围绕资产组织
采集器	执行指标采集的组件，与 Web 服务分离部署，具备自监控
降采样	把细粒度原始点聚合为粗粒度层，用于加速长周期查询
收敛	把同源的多条告警合并为一条，减少通知量
抑制	在特定条件下不产生某类告警，如上游故障时抑制下游衍生告警
风暴抑制	单位时间告警量超阈值时自动限流，防止通道被冲垮
基线	一组可核查的安全配置标准，如 CIS、等保技术要求
剧本	声明式的处置流程定义，含步骤、判据与回滚动作
机器指纹	由主机特征派生的短哈希，用于把授权绑定到指定主机
软性限额	超配额时仅禁止新增、不停止既有功能、不删除数据的限额方式

11.2 常见返回码

返回码	含义与处理
`FEATURE_NOT_LICENSED`	当前授权不含该功能。界面显示升级引导，接入方应按能力探测结果降级调用。
`QUOTA_EXCEEDED`	超出授权配额，仅拒绝本次新增操作，不影响既有数据与采集。
授权相关状态	接口返回当前授权状态与生效版本，可用于判断某能力是否可用。

11.3 版本对照

概念	取值	说明
软件版本	1.9.4	产品功能迭代版本；镜像 tag、安装包名与手册同号，由构建脚本从 package.json 注入
版本体系	V3.0	三档划分与闸门规则的版本，本手册所依据
交付形态	单轨 · 容器镜像	加固层已并入同一版本；镜像内置无头浏览器与中文字体，下载包约 240MB，离线 tar.gz 整包交付
近期能力增量	1.9.4	gNMI 厂商模板、GPU / DCGM 监控与纳管自动部署、PDF 内置浏览器、vSphere / 阿里云 ECS / K8s 云纳管、安装下载进度

CONTACT

图班 • cnzid
统一智能运维监控平台

技术支持
部署、升级、迁移与定制咨询

商务合作
企业版订阅 / OEM 贴牌 / 信创项目

扫码关注公众号

本手册所载功能、配额与版本归属以 V3.0 版本体系为准。产品持续演进，具体交付能力以实际版本说明与授权条款为准。
© 2026 cnzid · 图班.浑仪 · 功能说明手册 V3.0 · 1.9.4