Onsites AI

DEPLOYMENT GUIDE · EDITION 2.0

私有化部署指南

数据留在你自己的服务器上

一套镜像，两种网络：可连外网，也可以完全离线。模型与存储由你提供，我们不收这两项费用；安装、更新与支持由我们来做。



版本	发布	支持系统	文档编号
第 2.0 版	2026 年 11 月	Linux · Windows · macOS	ONS-SH-ZH-2.0

© 2026 Onsites AI · 本文档随产品迭代更新，最新版本见 onsitesai.com/cn/download.html

目录

CONTENTS

01 **什么时候该私有化** 三类组织、一张判断表、什么情况不必 …… 03

02 **架构与资源要求** 一套镜像、四段链路、离线与升级 …… 04

03 **四步上线** 评估、安装、接入、试点，各自的交付物 …… 05

04 **授权、费用与责任** 价格结构、与云端的差别、双方边界 …… 06

一页读懂

$20 /席位/月
年付 $240
10 席位起

24 小时
安装、配置与
迁移调试用时

60 天
完整产品试点
跑完再买授权

0 元
AI 模型与存储
我们都不收费

你的情况	建议形态	原因
客户资料、通话记录必须留在境内机房	私有化部署	出境即违规的场景，只有数据不出网才成立。
生产车间、军工或内网办公，完全不连外网	私有化部署（离线）	需要离线安装与离线包升级，不能靠联网激活。
必须用自建模型与自有存储	私有化部署	模型、密钥与磁盘都由自己提供，不买点数。
只是对隔离性有要求，仍希望我们运维	商业版 · 专属云	独占实例与驻留区域，但省掉机房与补丁。
常规团队，先把话接起来	云端 SaaS	15 秒开通，免费版即可验证流程。

这张表覆盖了绝大多数选型讨论；更细的判断见第 1 章。

文档用途

供 IT、安全与采购评估私有化部署的可行性、资源与费用，可直接作为立项附件。

价格口径

美元（USD），年付。最终报价以合同与订单为准；调整会同步官网与本文档。

不包含

本指南不含客户侧网络与硬件的采购实施，也不替代等保、行业合规的第三方测评。

版本与反馈

第 2.0 版，2026 年 11 月。评估问题请发送至 support@onsitesai.com。

01

WHEN TO SELF-HOST

三类组织，公有云解决不了的问题

私有化不是「更贵的 SaaS」，而是给那些数据不能出网、或必须自己掌握 AI 与存储的组织准备的。判断标准很简单：不是预算问题，而是数据能不能离开你的网络。

1.1 三类典型场景

场景一

数据不能出网

政府、银行、医疗与政策严格的子公司：客户资料、通话记录与销售单据必须留在境内、留在自己的机房，出境即为违规。

场景二

网络物理隔离

生产车间、军工与内网办公环境完全不连外网。软件要能离线运行，也要能用离线包升级，而不是「先连网激活一下」。

场景三

自建 AI 与存储

模型、密钥与文件存储都由自己提供：不买点数、不做用量计量，默认情况下客户数据不离开你的网络。

1.2 一条主线：数据在哪

问题 01

数据能否离开你的网络?

不能 → 私有化。能，但需要明确区域与隔离 → 专属云。都可以 → 云端 SaaS。

问题 02

是否必须离线运行?

是 → 私有化 + 离线包升级。否 → 三种形态都可以，按运维成本选。

问题 03

模型与密钥谁提供?

自己提供 → 私有化，我们不收 AI 与存储费用。由平台统一调用 → 云端或专属云。

问题 04

谁来做运维?

我们托管 → 云端 / 专属云。你的机房团队 + 我们的支持 → 私有化。

图 1-1　四个问题走完，形态基本就定了。任何一题答「必须自己掌握」，答案就收敛到私有化。

1.3 什么情况不必私有化

- **只是担心「数据安全」但没有硬性约束。**云端同样有 TLS 1.3、静态加密、席位权限、工作区隔离与审计记录，先看第 12 章（产品手册）再做判断。
- **团队没有机房与运维力量。**私有化需要有人管服务器、补丁与备份；没有的话，专属云是更省事的折中——实例独享、区域可选，仍由我们运维。
- **想先验证流程。**云端免费版就能跑通「一个队列接所有渠道」，验证完再谈部署形态，避免一次性投入。

> **一条经验**　真正需要私有化的组织，通常在第一次沟通时就会提到「出境」「等保」「内网」这三个词。如果一个都没提到，先从云端或专属云开始，成本与风险都更低。

02

ARCHITECTURE

一套镜像，四段链路

整套系统只有一个镜像：没有需要单独维护的中间件，也没有厂商绑定。云主机与内网服务器装的是同一个镜像，因此升级、备份与排障的方式完全一致。

2.1 部署拓扑



图 2-1 四段链路。中间两层完全在你的网络内；最后一层标注为可选——完全隔离的环境里，模型端点可以在内网，渠道出口可以只保留邮件。

2.2 资源清单

项目	要求
支持系统	Linux · Windows · macOS
最低配置	1 台虚拟机 · 4 核 · 8 GB 内存 · 3 Mbps 带宽
网络	可连外网，也可完全离线；离线环境使用离线包升级
AI 模型	你的 API 密钥、内部模型或物理隔离端点
存储	你自己的磁盘，容量由你管理；聊天与邮件文本占用极小，附件是主要消耗
备份	备份脚本随授权提供，按你的保留策略执行

2.3 离线与升级

- **离线安装。**安装包与依赖一起带走，不需要在部署时联网下载任何组件。
- **离线升级。**新版本以离线包交付，导入后按引导完成升级；不要求「先连网激活一下」。
- **排障方式一致。**云端与内网用同一套日志与健康检查，出问题时我们看的是同一份信息。
- **不发放自助安装包。**部署由我们完成——一次装错拖累的是生产环境。

资源怎么估 席位在 50 人以内、附件不堆到 TB 级时，4 核 8 GB 的单机足够；席位数、并发会话与附件体量是三个主要变量。拿不准就把预期规模告诉我们，评估阶段会给一份可直接采购的清单。

03

ROLLOUT

四步上线，每一步都有交付物

私有化的风险不在软件，而在「装完之后没人知道算不算装好了」。所以四步里每一步都约定交付物：评估给清单，安装给可用环境，接入给联调记录，试点给一份可以拿去立项的结论。

3.1 四个步骤

STEP 01

环境评估

确认架构、数据驻留与隔离要求，给出可直接采购的资源清单。

STEP 02

安装部署

支付一次性 $299 后由我们完成安装、配置与数据迁移调试，约 24 小时。

STEP 03

接入 AI 与存储

填入自有模型密钥或内网端点，指定磁盘位置；这两项我们不收费。

STEP 04

试点与采购

完整产品跑 60 天，跑完再按席位购买授权，不必一开始承担全年费用。

图 3-1　四步上线。第三步与第四步的顺序可以颠倒：也可以先试用、再接入模型。

3.2 交付物与时间

阶段	用时	你提供	我们交付
环境评估	1–3 个工作日	网络与隔离要求、预期规模	资源清单、架构确认书、实施排期
安装部署	约 24 小时	服务器、域名与证书	可登录的环境、健康检查报告
接入 AI 与存储	半天	模型端点、密钥、磁盘路径	联调记录、调用连通性测试
试点	60 天	参与人员与真实业务	使用报告、容量建议、授权报价

3.3 试点期建议关注的三件事

- **先接一个渠道。**网站聊天或主邮箱，跑通「客户进来 → 有人接 → 记录留下」，再开第二个渠道。
- **把单据跑一单。**报价 → 订单 → 合同 → 发票走一遍真实业务，确认编号、税率与审批能被复核。
- **给 AI 设上限。**先把 Copilot 打开，Harness 与 Desk 按需；软性上限跑两周，再决定是否改硬性上限。

> **试点不等于压测**　60 天试点用来验证流程与接受度，不是性能压测。如果预期峰值并发很高（例如大促期间的在线客服），请在评估阶段就说明，我们会按峰值给资源建议。

04

LICENSING

授权、费用与责任边界

席位价格里已经包含软件授权、软件更新与技术支持。AI 模型与存储由你自己提供，所以这条路线没有 AI 费，也没有存储费——这三件事分开看，账就很清楚。

4.1 价格结构

$20 /席位/月	$2,400 起	$299 一次	$0 AI/存储
年付 $240 / 席位 最少 10 个席位	10 席位年费 起步价	安装部署费 不抵扣授权费	模型与磁盘 由你提供

席位可随时增加，按剩余时间折算到续费日。安装部署费在试用开始前支付，由我们完成安装与调试，单独收取、不抵扣授权费。收款通过 2Checkout（Verifone）完成，可开据发票。

4.2 与云端的三处差别

对比项	云端（SaaS）	私有部署
开通时间	15 秒，无需安装	约 24 小时，由我们安装调试
数据位置	Onsites 云端；专属云可选区域	你的机房或你的私有云
离线运行	—	✓ 完全离线，用离线包升级
计费	$15 / 席位 / 月（年付 $180）	$20 / 席位 / 月（年付 $240），10 席位起
AI 与存储	点数按需购买，存储 $1/GB/月	由你提供，我们不收费
维护与升级	由我们托管运营	安装、更新与技术支持由我们提供

同一套软件、同一套功能。差异只有三处：数据放在哪、由谁维护、怎么付费。

4.3 双方的责任划分

事项	客户负责	我们负责
主机与网络	服务器、存储、网络与操作系统	软件安装、配置与升级支持
模型与密钥	提供端点与密钥，我们不持有	接入联调与连通性排查
备份	按脚本执行并保管备份介质	提供备份脚本与恢复步骤
安全事件	主机与网络层的监控与处置	软件层漏洞修复与通报

下一步 把环境说清楚（Docker / 虚拟机、物理隔离、私有云 VPC），我们会在一个工作日内给出可直接执行的部署答复；试点从支付 $299 安装部署费开始，跑完 60 天再决定是否购买授权。

评估申请：onsitesai.com/cn/self-host.html · 邮件：support@onsitesai.com