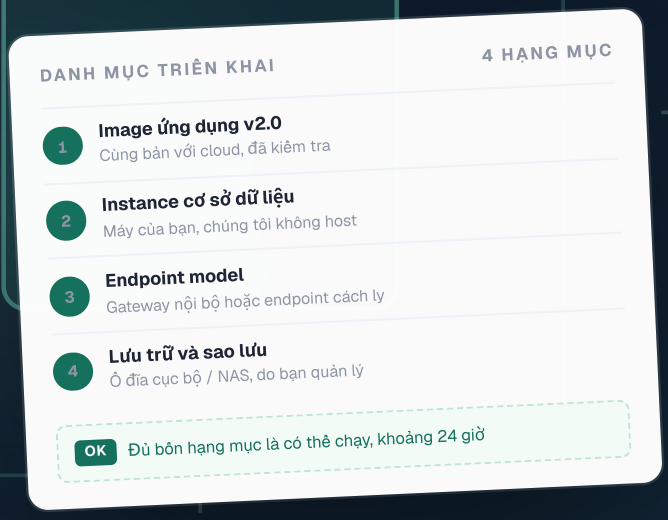
Onsites AI

DEPLOYMENT GUIDE · EDITION 2.0

Hướng dẫn triển khai self-hosted

Dữ liệu ở lại trên máy chủ của bạn

Một bộ image, hai kiểu mạng: có thể nối Internet hoặc hoàn toàn ngoại tuyến. Model và lưu trữ do bạn cung cấp, chúng tôi không thu phí hai hạng mục này; cài đặt, cập nhật và hỗ trợ do chúng tôi thực hiện.



Phiên bản
Bản 2.0

Phát hành
tháng 11 năm 2026

Hệ điều hành
Linux · Windows · macOS

Mã tài liệu
ONS-SH-VI-2.0

© 2026 Onsites AI · Tài liệu này được cập nhật theo sản phẩm; bản mới nhất tại onsitesai.com/vn/download.html

Mục lục

CONTENTS

01 **Khi nào nên self-hosted** Ba nhóm tổ chức, một bảng quyết định, khi nào không cần 03

02 **Kiến trúc và yêu cầu tài nguyên** Một bộ image, bốn chặng, ngoại tuyến và nâng cấp 04

03 **Bốn bước đưa vào dùng** Đánh giá, cài đặt, kết nối, dùng thử: đầu ra từng bước 05

04 **Bản quyền, chi phí và trách nhiệm** Cơ cấu giá, khác biệt so với cloud, ranh giới hai bên 06

Hiểu trong một trang

$20 /ghế/tháng
Trả năm $240 từ 10 ghế

24 giờ
Cài đặt, cấu hình và hiệu chỉnh di chuyển

60 ngày
Dùng thử trọn sản phẩm chạy xong mới mua

0 USD
Model AI và lưu trữ chúng tôi không thu phí

Tình huống của bạn	Hình thức đề xuất	Lý do
Hồ sơ khách hàng, ghi âm cuộc gọi phải lưu tại máy chủ trong nước	Self-hosted	Khi dữ liệu ra nước ngoài là vi phạm, chỉ dữ liệu không rời mạng mới đáp ứng.
Nhà máy, quốc phòng hoặc văn phòng mạng nội bộ, hoàn toàn không nối Internet	Self-hosted (ngoại tuyến)	Cần cài đặt ngoại tuyến và nâng cấp bằng gói ngoại tuyến, không thể kích hoạt qua mạng.
Bắt buộc dùng model tự xây và lưu trữ riêng	Self-hosted	Model, khóa và ổ đĩa đều tự cung cấp, không mua tín dụng AI.
Chỉ cần mức cách ly, vẫn muốn chúng tôi vận hành	Bản thương mại · cloud riêng	Instance riêng và vùng lưu trú, nhưng không phải lo máy chủ và bản vá.
Đội ngũ thông thường, cần nhận tin nhắn trước	Cloud SaaS	Mở trong 15 giây, bản miễn phí đủ để kiểm chứng quy trình.

Bảng này bao quát phần lớn các thảo luận chọn hình thức; tiêu chí chi tiết hơn ở Chương 1.

MỤC ĐÍCH

Giúp bộ phận IT, an ninh và mua sắm đánh giá tính khả thi, tài nguyên và chi phí của self-hosted; có thể dùng làm phụ lục trình duyệt dự án.

CƠ SỞ GIÁ

USD, trả theo năm. Báo giá cuối cùng theo hợp đồng và đơn hàng; mọi điều chỉnh sẽ cập nhật trên website và tài liệu này.

KHÔNG BAO GỒM

Tài liệu không gồm việc mua sắm, triển khai mạng và phần cứng phía khách hàng, cũng không thay thế đánh giá tuân thủ của bên thứ ba.

PHIÊN BẢN

Bản 2.0, tháng 11 năm 2026. Câu hỏi đánh giá xin gửi tới support@onsitesai.com.

01

WHEN TO SELF-HOST

Ba nhóm tổ chức mà public cloud không giải quyết được

Self-hosted không phải một SaaS đắt hơn, mà dành cho tổ chức có dữ liệu không được rời khỏi mạng, hoặc phải tự nắm model AI và lưu trữ. Tiêu chí rất đơn giản: không phải chuyện ngân sách, mà là dữ liệu có được rời khỏi mạng của bạn hay không.

1.1 Ba tình huống điển hình

Tình huống 1

Dữ liệu không được ra khỏi mạng

Cơ quan nhà nước, ngân hàng, y tế và các công ty con có chính sách chặt: hồ sơ khách hàng, ghi âm cuộc gọi và chứng từ bán hàng phải ở trong nước, trong máy chủ của chính mình; dữ liệu ra nước ngoài là vi phạm.

Tình huống 2

Cách ly vật lý về mạng

Nhà máy, quốc phòng và văn phòng mạng nội bộ hoàn toàn không nối Internet. Phần mềm phải chạy ngoại tuyến và nâng cấp bằng gói ngoại tuyến, không phải "kết nối mạng kích hoạt một lần".

Tình huống 3

Tự xây AI và lưu trữ

Model, khóa và lưu trữ tệp đều tự cung cấp: không mua tín dụng AI, không đo đếm mức dùng; mặc định dữ liệu khách hàng không rời khỏi mạng của bạn.

1.2 Một trục chính: dữ liệu ở đâu

Câu hỏi 01

Dữ liệu có được rời khỏi mạng của bạn?

Không → self-hosted. Có, nhưng cần vùng và mức cách ly rõ ràng → cloud riêng. Đều được → cloud SaaS.

Câu hỏi 02

Có bắt buộc chạy ngoại tuyến?

Có → self-hosted + nâng cấp bằng gói ngoại tuyến. Không → cả ba hình thức đều được, chọn theo chi phí vận hành.

Câu hỏi 03

Ai cung cấp model và khóa?

Tự cung cấp → self-hosted, chúng tôi không thu phí AI và lưu trữ. Nền tảng gọi tập trung → cloud hoặc cloud riêng.

Câu hỏi 04

Ai vận hành?

Chúng tôi vận hành → cloud / cloud riêng. Đội máy chủ của bạn + hỗ trợ của chúng tôi → self-hosted.

Hình 1-1. Đi hết bốn câu hỏi là gần như xác định được hình thức. Nếu bất kỳ câu nào trả lời "phải tự nắm", đáp án thu về self-hosted.

1.3 Khi nào không cần self-hosted

- **Chỉ lo "an toàn dữ liệu" mà không có ràng buộc cứng.** Cloud cũng có TLS 1.3, mã hóa dữ liệu tĩnh, phân quyền theo ghế, cách ly không gian làm việc và nhật ký kiểm toán; hãy xem Chương 12 (hướng dẫn sản phẩm) trước khi quyết định.
- **Đội ngũ không có máy chủ và nhân lực vận hành.** Self-hosted cần người quản máy chủ, bản vá và sao lưu; nếu không, cloud riêng là giải pháp dung hòa ít việc hơn — instance riêng, chọn được vùng, vẫn do chúng tôi vận hành.
- **Muốn kiểm chứng quy trình trước.** Bản miễn phí trên cloud đủ để chạy thông "một hàng đợi nhận mọi kênh", xong mới bàn hình thức triển khai, tránh đầu tư một lần.

> **Một kinh nghiệm:** Tổ chức thực sự cần self-hosted thường nhắc ngay trong lần trao đổi đầu ba từ: "ra nước ngoài", "tuân thủ", "mạng nội bộ". Nếu không từ nào xuất hiện, hãy bắt đầu từ cloud hoặc cloud riêng — chi phí và rủi ro đều thấp hơn.

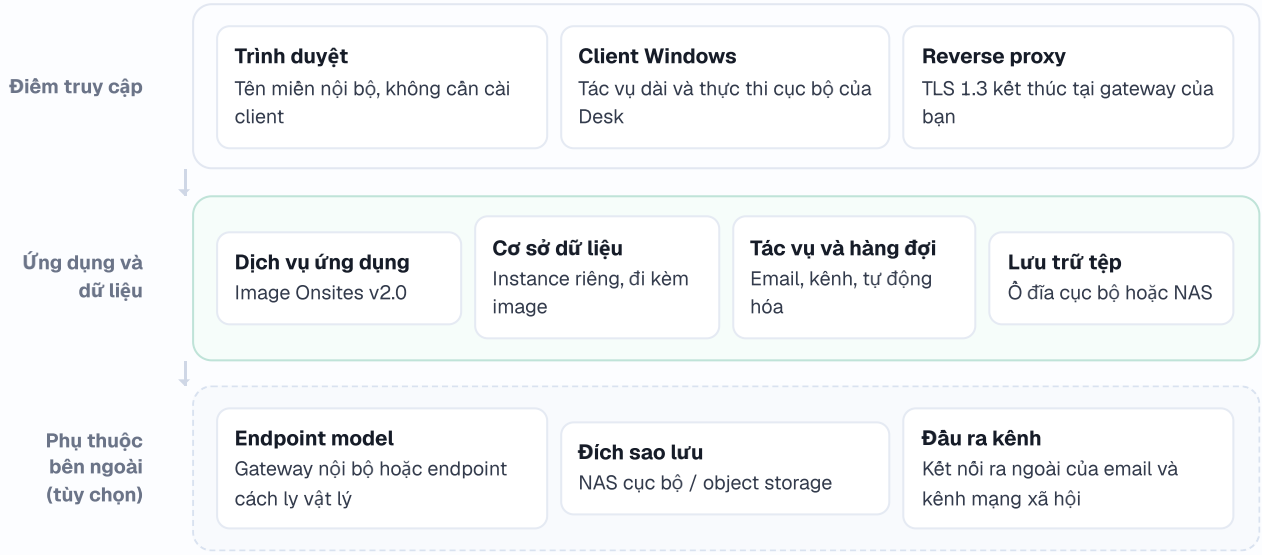
02

ARCHITECTURE

Một bộ image, bốn chặng

Toàn bộ hệ thống chỉ có một image: không middleware phải bảo trì riêng, không bị khóa vào nhà cung cấp. Máy chủ cloud và máy chủ nội bộ cài cùng image đó, nên nâng cấp, sao lưu và xử lý sự cố giống hệt nhau.

2.1 Sơ đồ triển khai



Hình 2-1. Bốn chặng. Hai lớp giữa nằm hoàn toàn trong mạng của bạn; lớp cuối là tùy chọn.

2.2 Danh mục tài nguyên

Hạng mục	Yêu cầu
Hệ điều hành	Linux · Windows · macOS
Cấu hình tối thiểu	1 máy ảo · 4 nhân · 8 GB RAM · băng thông 3 Mbps
Mạng	Nối Internet hoặc hoàn toàn ngoại tuyến; ngoại tuyến dùng gói nâng cấp ngoại tuyến
Model AI	Khóa API của bạn, model nội bộ hoặc endpoint cách ly vật lý
Lưu trữ	Ổ đĩa của bạn, dung lượng do bạn quản lý; chat và email chiếm rất ít, chủ yếu là tệp đính kèm
Sao lưu	Script sao lưu đi kèm giấy phép, chạy theo chính sách lưu giữ của bạn

2.3 Ngoại tuyến và nâng cấp

- **Cài đặt ngoại tuyến.** Bộ cài mang theo cả thành phần phụ thuộc, không cần tải gì từ mạng khi triển khai.
- **Cập nhật ngoại tuyến.** Bản mới giao bằng gói ngoại tuyến, nhập vào rồi làm theo hướng dẫn; không yêu cầu “kết nối mạng kích hoạt một lần”.
- **Cách xử lý sự cố giống nhau.** Cloud và mạng nội bộ dùng cùng bộ log và kiểm tra sức khỏe; khi có vấn đề, chúng tôi nhìn cùng một nguồn thông tin.
- **Không phát hành bộ cài tự phục vụ.** Việc triển khai do chúng tôi làm — cài sai một lần là hỏng production.

> **Ước lượng tài nguyên:** Với dưới 50 ghế và tệp đính kèm chưa tới mức TB, một máy 4 nhân 8 GB là đủ; số ghế, phiên đồng thời và khối lượng tệp là ba biến chính. Chưa chắc thì cho chúng tôi biết quy mô, danh mục mua sắm sẽ có ngay.

03

ROLLOUT

Bốn bước đưa vào dùng, mỗi bước đều có đầu ra

Rủi ro của self-hosted không nằm ở phần mềm mà ở chỗ “cài xong rồi không ai biết đã ổn hay chưa”. Vì vậy mỗi bước đều quy định đầu ra: đánh giá giao danh mục, cài đặt giao môi trường dùng được, kết nối giao biên bản chạy thử, dùng thử giao kết luận có thể dùng để trình duyệt dự án.

3.1 Bốn bước

BƯỚC 01

Đánh giá môi trường

Xác nhận kiến trúc, yêu cầu lưu trú dữ liệu và cách ly, đưa ra danh mục tài nguyên mua sắm được ngay.

BƯỚC 02

Cài đặt triển khai

Thanh toán $299 một lần, chúng tôi cài đặt, cấu hình và hiệu chỉnh di chuyển dữ liệu, khoảng 24 giờ.

BƯỚC 03

Kết nối AI và lưu trữ

Nhập khóa model riêng hoặc endpoint nội bộ, chỉ định vị trí ổ đĩa; hai hạng mục này không thu phí.

BƯỚC 04

Dùng thử và mua

Chạy trọn sản phẩm 60 ngày, xong mới mua giấy phép theo ghế, không phải chịu chi phí cả năm ngay từ đầu.

Hình 3-1. Bốn bước đưa vào dùng. Bước ba và bước bốn có thể đảo thứ tự: cũng có thể dùng thử trước rồi kết nối model.

3.2 Đầu ra và thời gian

Giai đoạn	Thời gian	Bạn cung cấp	Chúng tôi bàn giao
Đánh giá môi trường	1–3 ngày làm việc	Yêu cầu mạng và cách ly, quy mô dự kiến	Danh mục tài nguyên, biên bản xác nhận kiến trúc, lịch triển khai
Cài đặt triển khai	khoảng 24 giờ	Máy chủ, tên miền và chứng chỉ	Môi trường đăng nhập được, báo cáo kiểm tra sức khỏe
Kết nối AI và lưu trữ	nửa ngày	Endpoint model, khóa, đường dẫn ổ đĩa	Biên bản chạy thử, kiểm tra kết nối gọi model
Dùng thử	60 ngày	Nhân sự tham gia và nghiệp vụ thực	Báo cáo sử dụng, đề xuất dung lượng, báo giá giấy phép

3.3 Ba điều nên theo dõi khi dùng thử

- **Kết nối một kênh trước.** Chat website hoặc hộp thư chính, chạy thông “khách hàng vào → có người nhận → để lại ghi nhận”, rồi mới mở kênh thứ hai.
- **Chạy thử một chứng từ.** Báo giá → đơn hàng → hợp đồng → hóa đơn qua một nghiệp vụ thực, xác nhận số chứng từ, thuế suất và phê duyệt có thể kiểm tra lại.
- **Đặt giới hạn cho AI.** Mở Copilot trước, Harness và Desk tùy nhu cầu; chạy giới hạn mềm hai tuần rồi mới quyết định chuyển sang giới hạn cứng.

Dùng thử không phải kiểm thử tải: 60 ngày dùng thử để kiểm chứng quy trình và mức độ chấp nhận, không phải kiểm thử hiệu năng. Nếu dự kiến đồng thời đỉnh cao (ví dụ hỗ trợ trò chuyện mùa cao điểm), hãy nêu ở giai đoạn đánh giá, chúng tôi sẽ đề xuất tài nguyên theo mức đỉnh.

04

LICENSING

Bản quyền, chi phí và ranh giới trách nhiệm

Giá theo ghế đã bao gồm giấy phép phần mềm, cập nhật phần mềm và hỗ trợ kỹ thuật. Model AI và lưu trữ do bạn tự cung cấp, nên hướng này không có phí AI và cũng không có phí lưu trữ — tách ba khoản này ra thì rất rõ ràng.

4.1 Cơ cấu giá

$20 /ghế/tháng
Trả năm $240 / ghế
tối thiểu 10 ghế

$2,400 từ
Phí năm cho 10 ghế
giá khởi điểm

$299 một lần
Phí cài đặt triển khai
không trừ vào phí giấy
phép

$0 AI/lưu trữ
Model và ổ đĩa
do bạn cung cấp

Số ghế có thể tăng bất cứ lúc nào, tính theo thời gian còn lại tới ngày gia hạn. Phí cài đặt triển khai thanh toán trước khi bắt đầu dùng thử, do chúng tôi cài đặt và hiệu chỉnh, thu riêng và không trừ vào phí giấy phép. Thanh toán qua 2Checkout (Verifone), có thể xuất hóa đơn.

4.2 Ba khác biệt so với cloud

Hạng mục	Cloud (SaaS)	Self-hosted
Thời gian mở	15 giây, không cần cài đặt	Khoảng 24 giờ, chúng tôi cài đặt và hiệu chỉnh
Vị trí dữ liệu	Cloud Onsites; cloud riêng chọn được vùng	Máy chủ của bạn hoặc private cloud của bạn
Chạy ngoại tuyến	—	✓ Hoàn toàn ngoại tuyến, nâng cấp bằng gói ngoại tuyến
Cách tính phí	$15 / ghế / tháng (trả năm $180)	$20 / ghế / tháng (trả năm $240), từ 10 ghế
AI và lưu trữ	Tín dụng AI mua theo nhu cầu, lưu trữ $1/GB/tháng	Do bạn cung cấp, chúng tôi không thu phí
Bảo trì và nâng cấp	Chúng tôi vận hành	Cài đặt, cập nhật và hỗ trợ kỹ thuật do chúng tôi cung cấp

Cùng một phần mềm, cùng một bộ tính năng. Chỉ ba khác biệt: dữ liệu đặt ở đâu, ai bảo trì, trả tiền thế nào.

4.3 Phân chia trách nhiệm hai bên

Nội dung	Khách hàng	Onsites
Máy chủ và mạng	Máy chủ, lưu trữ, mạng và hệ điều hành	Cài đặt, cấu hình và hỗ trợ nâng cấp phần mềm
Model và khóa	Cung cấp endpoint và khóa, chúng tôi không nắm giữ	Chạy thử kết nối và kiểm tra khả năng kết nối
Sao lưu	Chạy theo script và bảo quản phương tiện sao lưu	Cung cấp script sao lưu và các bước phục hồi
Sự cố an ninh	Giám sát và xử lý lớp máy chủ, mạng	Vá lỗ hổng lớp phần mềm và thông báo

Bước tiếp theo: Hãy mô tả rõ môi trường (Docker / máy ảo, cách ly vật lý, VPC cloud riêng), chúng tôi sẽ trả lời triển khai có thể thực hiện ngay trong một ngày làm việc; dùng thử bắt đầu từ việc thanh toán phí cài đặt $299, chạy hết 60 ngày rồi mới quyết định mua giấy phép.
Đăng ký đánh giá: onsitesai.com/vn/self-host.html · Email: support@onsitesai.com