

2025 年|電子書

必讀指南：

掌握 AI 與雲原生部署之路

在混合多雲環境中，隨處輕鬆執行 AI 與雲原生應用程式

NUTANIX



目錄

Nutanix 如何解決 AI 與雲原生挑戰

GPT-in-a-Box	4
Nutanix Enterprise AI	4

更快速的程式碼與內容產出

考量因素.....	5
挑戰.....	5
Nutanix 如何提供協助.....	6
使用案例：政府機關處理文件分析.....	6

提升客戶體驗

考量因素.....	7
挑戰.....	7
Nutanix 如何提供協助.....	8
使用案例：衛星公司使用 GPT-in-a-Box 打造聊天機器人與 AI 助理	8

利用 AI 加強安全性

考量因素.....	9
挑戰.....	9
Nutanix 如何提供協助.....	10

加速部署雲原生應用程式

考量因素.....	11
挑戰.....	11
Nutanix 如何提供協助.....	11
使用案例：以 Nutanix 一站式打造雲原生環境	11

關鍵功能檢核清單

更快速的程式碼與內容產出.....	12
提升客戶體驗.....	12
利用 AI 加強安全性	12
加速部署雲原生應用程式.....	12

立即開始



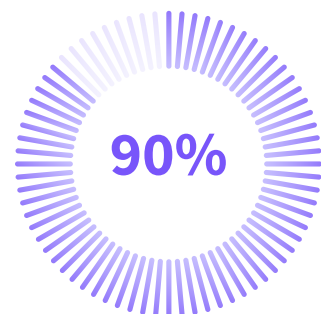
簡介

全球各地的企業正紛紛推動 AI 計畫，重點放在加速程式碼與內容產出、提升客戶體驗、強化安全防護，並加速部署雲原生應用程式的應用情境。

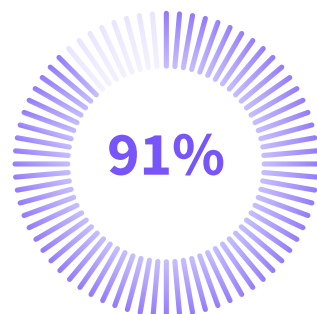
然而，許多組織仍在思考該從何處著手。為協助你順利前行，在規劃 AI 策略時需要克服兩項重大挑戰：

可能的使用案例過多：

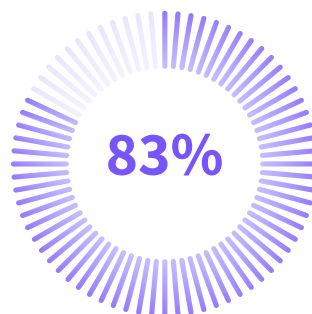
隨著 AI 使用情境大量湧現，往往難以聚焦於最能帶來最佳商業成果的項目。



將 AI 視為優先事項



同意他們的 IT 基礎架構
需要改善以支援 AI



計畫增加對邊緣策略的
投資以支援 AI

基礎架構複雜度：

選擇合適的 AI 基礎架構並支援 AI 應用程式於生產環境執行，是一項複雜決策。

要部署在地端、雲端或邊緣？需要哪些加速硬體？又該如何管理資料與 AI 生命週期？

雲原生環境中的 AI 應用程式通常以容器執行。習慣以虛擬機器作業的團隊，往往會發現設定與管理這類環境並不容易。

由於部署與資料模型不同、生態系演進快速，加上人才與經驗相對稀缺，挑戰更為明顯。

本電子書旨在協助你篩選合適的使用案例，並成功部署能真正推動業務成效的 AI 與其他雲原生應用程式。

內容探討四大領域的使用案例：

- **程式碼與內容產出：**導入 GenAI 以提升生產力。
- **客戶體驗：**為客戶提供更好、更個人化的體驗。
- **安全性：**保護企業、員工與客戶的安全與隱私。
- **雲原生應用程式：**使用最新雲原生工具開發並執行 AI 與其他應用程式。

本電子書包含多個真實案例，展示企業如何在各領域取得成功。不過在此之前，我們將先說明 Nutanix 採用的關鍵技術解決方案，協助你更瞭解在雲原生環境執行 AI 應用程式時，我們所解決的挑戰。

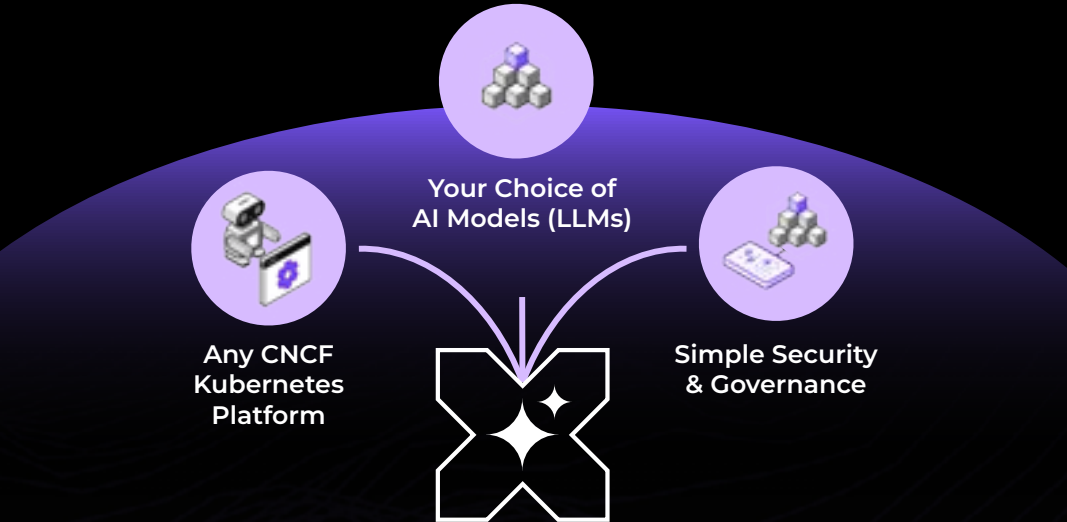
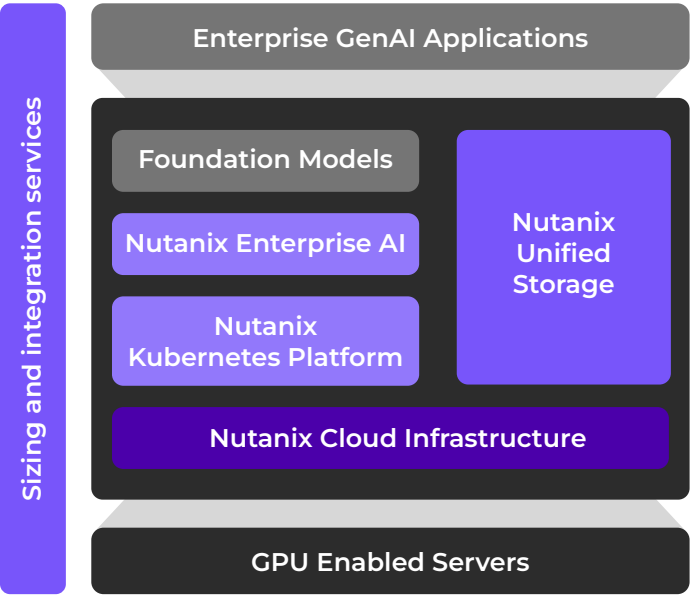
Nutanix 如何解決 AI 與雲原生挑戰

GPT-in-a-Box

Nutanix 推出的 GPT-in-a-Box 是軟體定義的統包式解決方案，旨在將 AI 模型無縫整合至你的作業中，同時讓資料與應用程式仍由你掌控。

優勢包括：

- **全 AI 堆疊**：使用主流開放原始碼 AI 框架部署精選的大型語言模型 (LLM) 組合。
- **隨處交付**：從資料中心到邊緣，再到公有雲。
- **內建資料服務**：在最佳化儲存上微調並執行 LLM，同時維持對資料與應用程式的掌控。
- **邊緣推論**：在邊緣據點部署執行 AI 推論模型所需的一切。精巧的佔用空間可克服場地限制並協助控管成本。
- **降低整體持有成本 (TCO)**：透過自動化、動態資源分配與整合，最佳化基礎架構成本以降低 TCO。
- **以立即可用的 AI 模型與 LLM 快速部署並靈活調整**：經驗證的 AI 模型可從邊緣到雲端快速部署，加速實現價值。
- **為 AI 應用程式部署並營運安全 API**：輕鬆建立安全的角色型 API，串接應用程式與 AI 模型。



Nutanix Enterprise AI

更快速的程式碼與內容產出

要開始導入 AI，請先找出能精簡業務流程並提升內部團隊效率的使用案例。常見例子包括：

- **內容輔助：**部署 AI 工具，在更短時間內產出更好的內容。
- **程式碼輔助：**以最新 AI 工具提升軟體團隊生產力。
- **智慧型文件處理：**運用自然語言處理自文件中擷取、分類並解讀資訊，加速業務流程。

許多組織因成本、安全與合規考量，尤其是受到法規嚴格規範的產業及公共部門，無法或不傾向使用 Microsoft、OpenAI 或 Google 等供應商的共用型雲端服務。但若以你可掌控的私有基礎架構支援程式碼與內容產出情境，你就能提供員工所需能力，同時避免善意使用者將敏感或專有資料上傳至公有 AI 服務所帶來的風險。

考量因素

程式碼與內容產出類的使用案例通常仰賴預先訓練的 GenAI 模型。若要開始使用，你需要部署全 AI 堆疊，並使用主流開放原始碼 AI 框架部署精選的大型語言模型 (LLM) 組合。

- 針對 Meta、Microsoft、Mistral 與 Google 等機構最新基礎模型的推論需求。Nutanix 已與開放原始碼模型的重要資源 Hugging Face 合作。

你可能還需要支援：

- 使用公司資料對所選模型進行微調 (深度學習的一種)。
- 建置檢索增強生成 (RAG)，讓模型可存取以資料庫整理的特定內部資料。

此 [Nutanix 驗證設計](#) 可協助你快速上手微調或建置 RAG。

挑戰

程式碼與內容產出使用案例常見挑戰包括：

- **複雜性：**你可能需要在所有營運據點部署 AI，包含資料中心、公有雲與邊緣。這些環境之間的技術差異會提高營運複雜性。
- **專用硬體：**你需要專用硬體來加速推論，例如 NVIDIA 的 GPU 或針對推論最佳化的 CPU，並支援 AI 訓練 (GPU)。由於此類硬體昂貴且不易取得，因此必須維持高使用率。
- **AI 生命週期管理：**由於 AI 應用程式及其所依賴的模型持續演進，通常偏好以容器與 Kubernetes 來執行與擴展 AI 模型，可帶來更快的部署與更新、高可用性與更有效的資源使用。然而，若容器與 Kubernetes 對組織而言仍屬新事物，可能出現技能落差，也可能難以找到具備雲原生技能的人才。

Nutanix 如何提供協助

Nutanix 透過安全、可擴展且具成本效益的平台，以更短時間與更少投入提供員工所需的 AI 工具。透過 Nutanix，你可在混合多雲環境中加速 AI 專案，同時維持對資料的掌控。

Nutanix 的優勢：

Nutanix 可解決程式碼與內容產出使用案例的多項關鍵挑戰。

- **善用混合多雲：**隨時隨地執行最新 AI 模型，無需在每個地點做出大幅營運調整。
- **有效利用專用硬體：**透過高效率的資源共享，讓昂貴的 GPU 與其他專用硬體發揮更高效益。
- **善用容器與 Kubernetes：**透過 GPT-in-a-Box 簡化 Kubernetes，Nutanix 讓你可使用偏好的工具，輕鬆踏上雲原生之路。同時，NAI 也讓你能在自己選擇的基礎架構與 Kubernetes 上輕鬆部署 AI 模型。

使用案例：政府機關處理文件分析

當某政府機關需要縮短在包含數百萬份、24 種語言文件的資料庫中尋找立法資料所耗費的時間時，他們選擇了 Nutanix。GPT-in-a-Box 讓該機關的 AI 解決方案更易於部署，並免除從頭開始建置的需要。

除了搭配 NVIDIA GPU 的 GPT-in-a-Box 之外，關鍵解決方案元素還包括：

- DocsGPT：AI 驅動的聊天機器人，可提升文件查詢類使用案例的使用者體驗。
- Red Hat OpenShift：完整的 Red Hat Kubernetes 全堆疊解決方案，已在 Nutanix 平台完成驗證。
- 來自 Mistral 的 LLM。

提升客戶體驗

GenAI 可提升並個人化客戶體驗、提高滿意度並激發創新，同時提升營運效率並降低客服成本。

常見的客戶體驗使用案例包括：

- **客服聊天機器人：**以公司資料微調 LLM，打造可處理客戶詢問、即時回應並協助解決複雜問題的聊天機器人。
- **情緒分析：**分析客戶回饋、辨識趨勢，並在問題擴大前主動處理。
- **個人化：**分析大量客戶偏好、行為與互動資料，產生個人化推薦、電子郵件、社群貼文及其他量身訂做的內容。

考量因素

如上述例子，客戶體驗類使用案例通常會使用 GenAI 基礎模型，並搭配微調或 RAG（檢索增強生成）。

你應規劃提供以下基礎架構支援：

- **針對所選基礎模型**的推論需求。
- **使用公司資料**進行模型訓練與微調。
- **建置 RAG**，讓所選模型可存取以資料庫整理的特定內部資料。

由於資料量龐大，客戶體驗使用案例通常會增加儲存需求並提高資料管理複雜性。

挑戰

當你規劃客戶體驗使用案例時，可能會遇到以下情況：

- **混合多雲：**你可能需要在所有營運據點部署 AI，而不同環境的技術差異會增加營運複雜性。
- **專用硬體：**需要 NVIDIA GPU 或針對推論最佳化的 CPU 來加速推論並支援 AI 訓練 (GPU)。
- **容器：**通常偏好以容器與 Kubernetes 來執行與擴展 AI 模型，以確保高可用性與有效的資源使用。

你也應預期資料管理挑戰將顯著增加。在 [2023 年度《Nutanix 企業 AI 現況報告》](#)指出，100% 的組織都需要額外技能，其中包含在 AI 生命週期中管理 AI 模型與資料集所需的能力。



Nutanix 如何提供協助

Nutanix 提供可提升客戶體驗的 AI 工具，例如更快解決問題，以及依客戶情境調整的回覆。透過 Nutanix，你可在混合多雲環境加速 AI 專案，同時簡化資料管理並維持對資料保護與隱私的掌控。

Nutanix GPT-in-a-Box 客戶體驗類使用案例通常仰賴最新的 GenAI 模型。Nutanix GPT-in-a-Box 可將 GenAI 與其他 AI 模型無縫整合至你的營運中，同時仍讓資料與應用程式由你掌控。

Nutanix Enterprise AI (NAI) 提供端點 API (包含 NVIDIA NIM 與 Hugging Face)，讓你可在地端與雲端輕鬆部署多種 GenAI 模型。你可將 NAI 部署在 Nutanix 基礎架構上，或部署在你選擇的基礎架構上。

完整的資料服務

在地端、雲端與邊緣環境中，效能與資料管理可能出現極大差異。Nutanix 透過簡化儲存營運，並提供建置與部署 AI 應用程式與服務所需的速度與彈性，克服這項挑戰。

使用案例：衛星公司使用 GPT-in-a-Box 打造聊天機器人與 AI 助理

一家快速成長的衛星營運商在 160 多個國家提供多任務衛星服務。該公司以提升生產力、效率、安全性與決策品質為 AI 策略重點，運用由 GPT-in-a-Box 驅動的 Nutanix AI 解決方案打造聊天機器人與 AI 助理。其中一個早期方案實現「對話查文件」(chat-to-documents)，打造出對公司知識非常深入的 AI 助理。

Nutanix 在基礎架構佈建、軟體部署與持續支援方面扮演關鍵角色，確保我們的 AI 專案得以順利推進並持續落實。

營運長

利用 AI 加強安全性

隨著網路犯罪持續攀升，資安使用案例成為優先要務，許多企業正轉向以 AI 來偵測並回應惡意行為者。

AI 資安使用案例包括：

- **詐騙偵測：**分析歷史交易資料，建立可辨識交易模式異常並精準偵測詐騙企圖的 AI 模型。
- **威脅偵測：**運用 AI 分析大量資料，在異常偵測、威脅情報與弱點評估等面向辨識模式。
- **警示強化：**使用 AI 自動蒐集並分析資安警示的脈絡資訊，協助你排序優先順序、降低誤報並加速事件應變。
- **自動建立政策：**根據持續性的環境分析、威脅態勢與產業最佳實務，動態產生並更新安全政策。

考量因素

能滿足這些使用案例的 AI 能力，如今已被納入最新的安全與資料保護工具中。

像詐騙偵測這類使用案例，可能需要你取得或自行建立 AI 模型。這需要團隊投入更多心力，包括大量實驗、定期訓練與微調。

目前在多數資安使用案例中，GenAI 仍多扮演輔助角色。例如，GenAI 常被用來生成合成的詐騙交易樣本，以提升用於詐騙偵測模型訓練資料集中的有效信號。

針對詐騙偵測或其他內部開發的資安軟體，你需要一套可支援以下需求的基礎架構：

- **AI 模型的推論。**
- **使用企業資料**進行模型訓練與微調。

在地端執行的商用資安軟體可能有特定基礎架構需求需滿足；而 SaaS 解決方案則可能有特定頻寬需求。

由於資料量龐大，資安使用案例可能增加儲存需求並提高資料管理複雜性。

挑戰

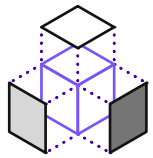
規劃資安使用案例時，應預期會遇到以下挑戰：

- **混合多雲：**你需要在所有營運據點部署 AI 驅動的資安軟體。
- **專用硬體：**需要——NVIDIA 的 GPU 或針對推論最佳化的 CPU——來加速推論並支援 AI 訓練 (GPU)。
- **容器：**通常偏好以容器與 Kubernetes 來執行與擴展 AI 模型，以確保高可用性與有效的資源使用。

你也應預期資料管理挑戰將顯著增加。許多組織缺乏在 AI 生命週期中管理 AI 模型與資料集所需的技能。

Nutanix 如何提供協助

Nutanix 協助你交付能主動防護業務的 AI 資安工具。透過 Nutanix，你可在混合多雲環境中加速 AI 資安專案，同時簡化資料管理。



完整的資料服務

資安使用案例屬於資料密集型，在進行即時威脅偵測時，你能否在資料中心、雲端與邊緣之間高效率地移動與管理資料，往往至關重要。

Nutanix 透過軟體定義資料服務簡化資料中心、雲端與邊緣的儲存營運，進而消除資料管理挑戰。

Nutanix 結合你所需的資料服務與你期待的企業級能力。我們免除分別管理檔案、區塊與物件儲存系統的需求，同時降低成本與複雜度性。



安全強化的平台

在 Nutanix，安全始於專為混合多雲架構打造的穩固軟體基礎。Nutanix 先以已完成安全強化的超融合軟體平台為起點，並在此基礎上持續延伸。

我們的功能可協助你快速偵測並回應可能導致資料遺失與業務中斷的威脅，進而強化整體安全態勢。你可以信賴 Nutanix 為 AI 與其他關鍵工作負載提供安全的基礎。



加速部署雲原生應用程式

企業 IT 團隊雖然在虛擬化方面經驗豐富，但對 Kubernetes、容器與微服務的細節可能缺乏經驗，甚至完全陌生。若要避免正式上線後產生營運障礙，成功地大規模部署容器化應用程式，你需要現代化你的方法與技能。

考量因素

雲原生技術可打造低耦合系統，使其具備韌性、可管理性與可觀測性。搭配強大的自動化後，工程師便能在最少投入、最低阻礙下，頻繁且可預期地進行具有高影響力的改變。

挑戰

設定與管理雲原生環境（包含儲存、網路、監控與 Kubernetes 生命週期管理）是一項艱鉅挑戰。首次部署雲原生者通常會在四個關鍵面向遇到挑戰：

- **部署和管理：**設計、部署與營運全堆疊解決方案需要大量時間與專業，尤其是在資料中心、雲端與邊緣方面。
- **專業人才：**相關專業供給不足。雲原生領域的許多方案採開放原始碼，導致營運支援更難取得。
- **生態系統：**雲原生生態系統成長快速且持續演進，因此必須找到能協助你辨別重點、去蕪存菁的技術夥伴。
- **資料：**容器執行個體是暫時性的，因此需要具備彈性，才能滿足檔案、物件與區塊儲存，以及各類資料庫的持久性儲存需求。

Nutanix 如何提供協助

Nutanix 可讓你在資料中心、跨多個雲端或邊緣位置，輕鬆部署一致的端對端生產就緒型雲原生環境。

隨時隨地執行

無論你在地端、公有雲、服務供應商資料中心或邊緣執行雲原生工作負載，Nutanix 都能在各環境提供一致的管理與安全防護。

Nutanix Kubernetes 平台

[Nutanix Kubernetes 平台](#) (NKP) 透過降低營運複雜度來簡化平台工程，同時不會將你綁定在單一 Kubernetes 選項上。與 Red Hat、Microsoft、Google Cloud 和 AWS 的策略合作夥伴關係，讓你能使用偏好的 Kubernetes 平台。

NKP 的優勢包括：

全方位的資料服務

- [適用於 Kubernetes 的 Nutanix 資料服務](#) (NDK)
- 韌性
- 緊密整合
- 輕鬆升級

使用案例：以 Nutanix 一站式打造雲原生環境

一家製藥公司希望把雲原生環境所需的一切進行整合並串聯。IT 團隊擔心自己是否有能力管理這個分散的環境，因為一旦出問題，很難取得支援與指引。

透過 Nutanix，IT 團隊打造出功能完整、可擴展且易於管理的雲原生環境。如今，團隊透過經 Nutanix 驗證的 Red Hat OpenShift Plus 解決方案所提供的虛擬化平台，不僅為其帶來韌性和成本效益，還省下大量時間與人力。

透過與 Nutanix 攜手合作，該公司獲得了值得信賴的合作夥伴分享最佳實務與藍圖，並在需要的時間與地點提供協助。

關鍵功能檢核清單

剛開始導入 AI 與雲原生應用程式的企業，可先從較直接的使用案例著手：程式碼與內容產出、客戶體驗、安全，以及雲原生使用案例。

更快速的程式碼與內容產出

程式碼和內容使用案例 (如輔助功能) 仰賴 GenAI 模型的私有化部署。關鍵功能包括：

支援推論的基礎架構，具備：

- ✓ 合適的加速器硬體
- ✓ 能在資料中心、雲端與邊緣各處彈性部署
- ✓ 大規模支援容器化應用程式的能力

Nutanix 以 GPT-in-a-Box 加速程式碼與內容使用案例；這是一套軟體定義的統包式解決方案，可將 GenAI 模型無縫整合至你的作業中。

提升客戶體驗

客戶體驗使用案例可個人化客戶體驗、提高滿意度、提升營運效率，並降低客服成本。關鍵功能包括：

- ✓ 支援推論的基礎架構，具備：
 - ✓ 合適的加速器硬體
 - ✓ 能在資料中心、雲端與邊緣各處彈性部署

✓ 支援 AI 模型訓練與微調的基礎架構

✓ 支援 RAG 的能力

✓ 在整個 AI 生命週期中管理 AI 模型與資料集所需的基礎架構與技能

✓ 大規模支援容器化應用程式

Nutanix 提供基礎架構，協助你在混合多雲環境中加速 AI 專案，同時簡化資料管理，並讓你維持對資料保護與隱私的掌控度。

利用 AI 加強安全性

企業愈來愈依賴 AI 來強化安全防護，資安使用案例也成為許多組織的優先要務。關鍵功能包括：

- ✔ 內部開發的資安軟體
- ✔ 支援推論的基礎架構，具備：
 - ✔ 合適的加速器硬體
 - ✔ 能在資料中心、雲端與邊緣各處彈性部署
- ✔ 支援 AI 模型訓練與微調的基礎架構
- ✔ 在整個 AI 生命週期中管理 AI 模型與資料集所需的基礎架構與技能
- ✔ 大規模支援容器化應用程式

在地端執行的商用資安軟體可能有特定基礎架構需求需滿足；而 SaaS 則可能有特定頻寬需求。

Nutanix 提供基礎架構，協助你在混合多雲環境中加速 AI 資安專案。[Nutanix Security Central](#) 使用機器學習演算法分析網路流量，並針對特定環境建議合適的政策。

加速部署雲原生應用程式

領先企業正採用雲原生開發方法，以加速數位服務交付、降低客戶互動摩擦，並為員工提供最新的資料驅動工具。關鍵功能包括：

- ✔ 能在資料中心、雲端與邊緣執行同一套雲原生堆疊
- ✔ 依特定應用程式需求選擇合適元件的彈性
- ✔ 可在雲原生環境中高效運作的應用程式感知資料服務
- ✔ 整合以簡化並加速雲原生應用程式堆疊的部署和管理，並依你的需求量身打造
- ✔ 能瞭解雲原生生態系統且支援雲原生堆疊的合作夥伴

Nutanix 讓你能在資料中心、雲端與邊緣交付並管理同一套端對端、生產就緒的雲原生環境。我們安全、具韌性且可自我修復的平台，可在任何位置簡化雲原生應用程式的規劃、部署與營運。

立即開始

Nutanix 提供你因應組織最關鍵使用案例所需的一切能力。透過免除整合各種異質硬體與軟體的需求，Nutanix 提供可擴展、彈性且易管理的 AI 平台，為你省下數個月的心力。

別讓任何事物阻礙你在 AI 與雲原生領域取得更大的成功。歡迎親自體驗 AI 與雲原生測試環境，親眼見證 Nutanix 的與眾不同。

開始試用 AI

歡迎試用雲原生服務

請造訪 nutanix.com/ai 以瞭解更多資訊。您你可以傳送電子郵件至 info@nutanix.com 聯絡 Nutanix，或前往 www.nutanix.com/demo 將你的需求傳送給我們，我們會為你精心客製專屬簡報。

NUTANIX

info@nutanix.com | www.nutanix.com | [@nutanix](https://twitter.com/nutanix)

©2026 Nutanix, Inc. 保留所有權利。Nutanix、Nutanix 標誌以及本文提及的所有產品和服務名稱均為 Nutanix, Inc. 的註冊商標或商標。其他國家的註冊商標或商標。此處提及的所有其他品牌名稱均僅供識別參考，且可能為其各自擁有者所屬商標。
AI-EssentialGuidetoAI-eBook-FY25Q2-v3_zh-TW_061726

