



NetApp AIPod Mini for ERAG - 基礎架構就緒檢查清單

NetApp artificial intelligence solutions

NetApp
February 12, 2026

目錄

NetApp AIPOD Mini for ERAG - 基礎架構就緒檢查清單	1
業務與使用案例準備度	1
技術與基礎架構需求	1
軟體與生態系統協調	1
資料治理與安全	1
營運與支援考量	2
商業與 GTM 協調一致	2

NetApp AI Pod Mini for ERAG - 基礎架構就緒檢查清單

本文檔概述了 NetApp AI Pod Mini for Enterprise RAG 的全面基礎設施準備清單，可作為部署前參考。

業務與使用案例準備度

- 該解決方案是否與業務線成果（例如生產力、客戶服務、法律、醫療保健、製造業、公共部門）一致？
- 您是否已估算 LLM 工作負載的首次令牌獲取時間 (TTFT) 和延遲需求？
- 您知道預期的並發 / 使用者負載嗎（例如，RAG 中每個 2 個工作節點 32 個並髮使用者）？
- 您是否已確定主要的 AI/GenAI 工作負載（RAG、inferencing、fine-tuning、部門 LLM、vector DB 整合）？
- 您是否正在評估基於 CPU 的 AI 方案（OPEA、Intel Xeon）與 GPU 方案在成本 / 效能平衡方面的優劣？

技術與基礎架構需求

- 您的資料管道是否已準備就緒（資料準備、ETL、安全導入向量資料庫）？
- 您是否需要高可用性、備援或 DR 功能？
- 您是否正在利用 Ubuntu Linux / Kubernetes / Red Hat OpenShift AI 技術堆疊的支援？
- 您是否已驗證網路效能（根據工作負載的不同，速度範圍為 25–100GbE）？
- 儲存是否已配置 NetApp ONTAP + Trident CSI 驅動程式以實現 Kubernetes 持久化？
- 至少需要 3 個運算節點 (2 個工作節點、1 個控制平面節點)，且大小是否正確？

軟體與生態系統協調

- 您的容器化應用程式是否與提供的 Kubernetes 和 Helm charts 相容？
- 計劃部署哪些向量資料庫（例如 Milvus、pgvector）？
- 您是否需要 OPEA（Open Platform for Enterprise AI）預先整合來實現檢索增強生成（RAG）？
- 您是否正在利用混合雲選項（Cloud Volumes ONTAP、FSxN、Anthos、Azure Arc）？
- 您是否需要合作夥伴 ISV 整合（ESRI、醫療保健 PACS、金融 / 法律 ISV）？

資料治理與安全

- 您是否已在 Kubernetes 中啟用角色型存取控制（RBAC）？
- 是否有資料保護和備份計劃（SnapMirror、SnapCenter、勒索軟體防護）？
- 您是否已規劃資料法規遵循需求（HIPAA、GDPR、FedRAMP、CJIS）？
- 您是否需要私有 AI 部署（實體隔離、內部部署、安全區域）？

營運與支援考量

- 管理者是否接受過 Kubernetes、Trident CSI 和 OPEA 技術棧部署的訓練/具備相關能力？
- 您是否需要多租戶支援（部門、SLED 機構、業務部門）？
- 是否有監控和可觀測性計劃（ONTAP System Manager、Cloud Insights、Prometheus/Grafana）？
- 誰將負責第 2 天營運（客戶 IT 部門、合作夥伴、託管服務提供者）？

商業與 GTM 協調一致

- 是否有分階段的路線圖（從部門級到企業級 AI 擴展）？
- 您是否有多年期的預測報告（TCO、ARR、利潤率提升）？
- 授權升級情境是否明確（向量資料庫、ISV 軟體、AI 維運工具）？
- 您是否考慮過合作夥伴獎勵措施（經銷商利潤、OEM/Intel 聯合出資）？
- 購買行為是否符合預算週期（CapEx 與 OpEx、消費模式）？
- 您是否有服務合作夥伴（Arrow、WWT、Presidio 等）負責規模評估和部署？

版權資訊

Copyright © 2026 NetApp, Inc. 版權所有。台灣印製。非經版權所有人事先書面同意，不得將本受版權保護文件的任何部分以任何形式或任何方法（圖形、電子或機械）重製，包括影印、錄影、錄音或儲存至電子檢索系統中。

由 NetApp 版權資料衍伸之軟體必須遵守下列授權和免責聲明：

此軟體以 NETAPP「原樣」提供，不含任何明示或暗示的擔保，包括但不限於有關適售性或特定目的適用性之擔保，特此聲明。於任何情況下，就任何已造成或基於任何理論上責任之直接性、間接性、附隨性、特殊性、懲罰性或衍生性損害（包括但不限於替代商品或服務之採購；使用、資料或利潤上的損失；或企業營運中斷），無論是在使用此軟體時以任何方式所產生的契約、嚴格責任或侵權行為（包括疏忽或其他）等方面，NetApp 概不負責，即使已被告知有前述損害存在之可能性亦然。

NetApp 保留隨時變更本文所述之任何產品的權利，恕不另行通知。NetApp 不承擔因使用本文所述之產品而產生的責任或義務，除非明確經過 NetApp 書面同意。使用或購買此產品並不會在依據任何專利權、商標權或任何其他 NetApp 智慧財產權的情況下轉讓授權。

本手冊所述之產品受到一項（含）以上的美國專利、國外專利或申請中專利所保障。

有限權利說明：政府機關的使用、複製或公開揭露須受 DFARS 252.227-7013（2014 年 2 月）和 FAR 52.227-19（2007 年 12 月）中的「技術資料權利 - 非商業項目」條款 (b)(3) 小段所述之限制。

此處所含屬於商業產品和 / 或商業服務（如 FAR 2.101 所定義）的資料均為 NetApp, Inc. 所有。根據本協議提供的所有 NetApp 技術資料和電腦軟體皆屬於商業性質，並且完全由私人出資開發。美國政府對於該資料具有非專屬、非轉讓、非轉授權、全球性、有限且不可撤銷的使用權限，僅限於美國政府為傳輸此資料所訂合約所允許之範圍，並基於履行該合約之目的方可使用。除非本文另有規定，否則未經 NetApp Inc. 事前書面許可，不得逕行使用、揭露、重製、修改、履行或展示該資料。美國政府授予國防部之許可權利，僅適用於 DFARS 條款 252.227-7015(b)（2014 年 2 月）所述權利。

商標資訊

NETAPP、NETAPP 標誌及 <http://www.netapp.com/TM> 所列之標章均為 NetApp, Inc. 的商標。文中所涉及的所有其他公司或產品名稱，均為其各自所有者的商標，不得侵犯。