



NetApp AIPod Mini for ERAG - 基础设施准备情况核对表

NetApp artificial intelligence solutions

NetApp
February 12, 2026

目录

NetApp AIPOD Mini for ERAG - 基础设施准备情况核对表	1
业务和用例就绪性	1
技术和基础设施要求	1
软件和生态系统协调	1
数据治理与安全	1
运营和支持注意事项	2
商业和 GTM 对齐	2

NetApp AI Pod Mini for ERAG - 基础设施准备情况 核对表

本文档概述了 NetApp AI Pod Mini for Enterprise RAG 的综合基础架构就绪性检查清单，作为部署前参考。

业务和用例就绪性

- 解决方案是否与业务线结果一致（例如，生产力、客户服务、法律、医疗保健、制造业、公共部门）？
- 您是否已经估计了 LLM 工作负载的首次令牌时间 (TTFT) 和延迟需求？
- 您是否知道预期的并发/用户负载（例如，对于 RAG，每个 2-worker 节点 32 个并发用户）？
- 您是否已确定主要 AI/GenAI 工作负载（RAG、推理、微调、部门 LLM、矢量数据库集成）？
- 您是否正在评估基于 CPU 的 AI 选项（OPEA、Intel Xeon）与 GPU 替代方案之间的成本/性能平衡？

技术和基础设施要求

- 您的数据管道准备好了吗（数据准备、ETL、安全提取到矢量数据库中）？
- 是否需要高可用性、冗余或 DR 功能？
- 您是否正在利用 Ubuntu Linux / Kubernetes / Red Hat OpenShift AI 堆栈支持？
- 您是否已验证网络性能（25-100GbE，具体取决于工作负载）？
- 是否使用 NetApp ONTAP + Trident CSI 驱动程序为 Kubernetes 持久化配置存储？
- 是否正确调整了至少 3 个计算节点（2 个工作节点，1 个控制平面）的大小？

软件和生态系统协调

- 您的容器化应用程序是否与提供的 Kubernetes & Helm 图表兼容？
- 计划部署哪些矢量数据库（例如 Milvus、pgvector）？
- 是否需要 OPEA（Open Platform for Enterprise AI）预集成以进行检索增强生成（RAG）？
- 您是否正在利用混合云选项（Cloud Volumes ONTAP、FSxN、Anthos、Azure Arc）？
- 是否需要合作伙伴 ISV 集成（ESRI、医疗保健 PACS、金融/法律 ISV）？

数据治理与安全

- 您是否在 Kubernetes 中启用了基于角色的访问控制 (RBAC)？
- 是否有数据保护和备份计划（SnapMirror、SnapCenter、勒索软件保护）？
- 您是否已确定数据合规需求（HIPAA、GDPR、FedRAMP、CJIS）？
- 是否需要私有 AI 部署（气隙、本地、安全飞地）？

运营和支持注意事项

- 管理员是否接受过 Kubernetes、Trident CSI 和 OPEA 堆栈部署方面的培训/启用？
- 是否需要为多租户（部门、SLED 机构、业务部门）提供支持？
- 是否有监控和可观察性计划（ONTAP System Manager、Cloud Insights、Prometheus/Grafana）？
- 谁将负责第 2 天的运营（客户 IT、合作伙伴、托管服务提供商）？

商业和 GTM 对齐

- 是否有分阶段的路线图（部门→企业范围的 AI 扩展）？
- 您是否有多年期备考预测（TCO、ARR、利润率提升）？
- 许可提升场景是否明确（矢量数据库、ISV 软件、AI 运维工具）？
- 您是否探索过合作伙伴奖励（分销商利润率、OEM/Intel 联合融资）？
- 采购是否与预算周期（CapEx 与 OpEx、消费模式）保持一致？
- 您是否有服务合作伙伴（Arrow、WWT、Presidio 等）进行规模调整和部署？

版权信息

版权所有 © 2026 NetApp, Inc.。保留所有权利。中国印刷。未经版权所有者事先书面许可，本档中受版权保护的任何部分不得以任何形式或通过任何手段（图片、电子或机械方式，包括影印、录音、录像或存储在电子检索系统中）进行复制。

从受版权保护的 NetApp 资料派生的软件受以下许可和免责声明的约束：

本软件由 NetApp 按“原样”提供，不含任何明示或暗示担保，包括但不限于适销性以及针对特定用途的适用性的隐含担保，特此声明不承担任何责任。在任何情况下，对于因使用本软件而以任何方式造成的任何直接性、间接性、偶然性、特殊性、惩罚性或后果性损失（包括但不限于购买替代商品或服务；使用、数据或利润方面的损失；或者业务中断），无论原因如何以及基于何种责任理论，无论出于合同、严格责任或侵权行为（包括疏忽或其他行为），NetApp 均不承担责任，即使已被告知存在上述损失的可能性。

NetApp 保留在不另行通知的情况下随时对本文档所述的任何产品进行更改的权利。除非 NetApp 以书面形式明确同意，否则 NetApp 不承担因使用本文档所述产品而产生的任何责任或义务。使用或购买本产品不表示获得 NetApp 的任何专利权、商标权或任何其他知识产权许可。

本手册中描述的产品可能受一项或多项美国专利、外国专利或正在申请的专利的保护。

有限权利说明：政府使用、复制或公开本文档受 DFARS 252.227-7013（2014 年 2 月）和 FAR 52.227-19（2007 年 12 月）中“技术数据权利 — 非商用”条款第 (b)(3) 条规定的限制条件的约束。

本文档中所含数据与商业产品和/或商业服务（定义见 FAR 2.101）相关，属于 NetApp, Inc. 的专有信息。根据本协议提供的所有 NetApp 技术数据和计算机软件具有商业性质，并完全由私人出资开发。美国政府对这些数据的使用权具有非排他性、全球性、受限且不可撤销的许可，该许可既不可转让，也不可再许可，但仅限在与交付数据所依据的美国政府合同有关且受合同支持的情况下使用。除本文档规定的情形外，未经 NetApp, Inc. 事先书面批准，不得使用、披露、复制、修改、操作或显示这些数据。美国政府对国防部的授权仅限于 DFARS 的第 252.227-7015(b)（2014 年 2 月）条款中明确的权利。

商标信息

NetApp、NetApp 标识和 <http://www.netapp.com/TM> 上所列的商标是 NetApp, Inc. 的商标。其他公司和产品名称可能是其各自所有者的商标。