



发行说明 GenAI

NetApp
October 06, 2025

This PDF was generated from <https://docs.netapp.com/zh-cn/workload-genai/whats-new.html> on October 06, 2025. Always check docs.netapp.com for the latest.

目录

- 发行说明 1
 - NetApp Workload Factory for GenAI 的新功能 1
 - 2025年10月5日 1
 - 2025年8月3日 1
 - 2025年6月29日 2
 - 2025年6月3日 2
 - 2025年5月4日 2
 - 2025年3月2日 3
 - 2025年2月2日 3
 - 2025年1月5日 4
 - 2024年12月1日 4
 - 2024年11月3日 5
 - 2024年9月29日 5
 - 2024年9月1日 6
 - 2024 年 8 月 4 日 6
 - 2024年7月7日 6

发行说明

NetApp Workload Factory for GenAI 的新功能

了解 Workload Factory 的生成式 AI 工作负载功能的新功能。

2025年10月5日

BlueXP workload factory现为**NetApp**工作负载工厂

BlueXP已重命名并重新设计，以更好地反映其在管理数据基础设施中的作用。因此，BlueXP workload factory 已重命名为NetApp工作负载工厂。

支持在适用于 **Amazon Q Business** 的**NetApp**连接器中添加通用 **NFS/SMB** 数据源

使用 Workload Factory API，您现在可以将数据源从通用 NFSv3、NFSv4 或 SMB 共享添加到NetApp Connector for Amazon Q Business。这使您能够包含存储在由Amazon FSx for NetApp ONTAP以外的文件系统托管的卷上的文件。

["为 Amazon Q Business 创建 NetApp 连接器"](#)

["向连接器添加数据源"](#)

知识库的高级聊天配置

您现在可以配置适用于知识库聊天模型的高级聊天设置，例如响应长度、温度、推理设置等。其中一些设置（例如最近时间和修改时间设置、高级检索设置和系统提示）仅可使用 Workload Factory API。

["创建GenAI知识库"](#)

现支持嵌入、聊天和重新排名模型的推理类型选择

如果您选择的嵌入、聊天或重新排名模型具有推理设置，您现在可以选择推理类型。这使您能够更好地根据您的需要调整聊天机器人的性能和资源需求。

["创建GenAI知识库"](#)

2025年8月3日

结构化数据结果的安全存储

如果聊天机器人查询结果包含结构化数据，GenAI 可以将结果存储在 Amazon S3 存储桶中。当这些结果存储在 S3 存储桶中时，您可以使用聊天会话中的下载链接下载它们。

["创建GenAI知识库"](#)

MCP 服务器可用性

NetApp现在为 GenAI 提供带有NetApp Workload Factory 的模型上下文协议 (MCP) 服务器。您可以在本地安装

服务器，以允许外部 MCP 客户端从 GenAI 知识库中发现和检索查询结果。

["NetApp Workload Factory GenAI MCP 服务器"](#)

2025年6月29日

支持托管在通用 **NFS/SMB** 文件系统上的数据源

您现在可以从通用 SMB 或 NFS 共享添加数据源。这使您能够包含存储在 Amazon FSx for NetApp ONTAP 以外的文件系统托管的卷上的文件。

["向知识库添加数据源"](#)

["向连接器添加数据源"](#)

2025年6月3日

跟踪器可用于监控和跟踪操作

GenAI 现已提供 Tracker 监控功能。您可以使用 Tracker 监控和跟踪待处理、正在进行和已完成操作的进度和状态，查看操作任务和子任务的详细信息，诊断任何问题或故障，编辑失败操作的参数，以及重试失败的操作。

["使用NetApp Workload Factory 中的 Tracker 监控工作负载操作"](#)

为知识库选择重排序模型

现在，您可以通过选择与知识库配合使用的特定重排序模型来提高重排序查询结果的相关性。GenAI支持 Cohere Rerank 和 Amazon Rerank 模型。

["创建GenAI知识库"](#)

2025年5月4日

支持 **NetApp Connector for Amazon Q Business**

此版本的 GenAI 引入了对 NetApp Connector for Amazon Q Business 的支持，使您能够为 Amazon Q Business 创建连接器。快速轻松地利用Amazon Q Business AI助理、与构建Amazon Brock的GenAI知识库相比、初始配置更少。

["为 Amazon Q Business 创建 NetApp 连接器"](#)

增强了聊天模式支持

现在、GenAI还支持在知识库中使用以下聊天模式：

- ["Mistral AI型号"](#)
- ["Amazon T人 文本模型"](#)
- ["Meta Llama模型"](#)
- ["JAMBA 1.5型号"](#)

- ["Cohere命令模型"](#)
- ["DeepSeek型号"](#)

GenAI支持Amazon Brock支持的每个提供商的型号：["Amazon Brock中支持的基础模型"](#)

["创建GenAI知识库"](#)

更新了权限术语

Workload Factory 用户界面和文档现在使用“只读”来指代读取权限，使用“读/写”来指代自动化权限。

2025年3月2日

嵌入式聊天机器人增强功能

现在、您可以将问题和回答直接复制到剪贴板、调整聊天窗口的大小并更改其标题。此外、聊天响应现在可以包括表、这些表也是可复制的。

["测试GenAI知识库"](#)

聊天响应引用支持

聊天回复现在包含引用、列出用于生成回复的文件和数据块。

["测试GenAI知识库"](#)

增强了文件类型支持

此版本的GenAI提供了增强的文件支持：

- 聊天模式改进了CSV支持。这样、在从CSV文件查询数据时、可以做出更有用的响应。
- 现在、GenAI可以从数据源中安装Apache Parquet文件。
- GenAI现在支持载入包含图像的Microsoft Word DOCX文件。对DOCX文档中嵌入的图像进行扫描、对知识库查询的响应中包含嵌入图像中的文本洞察信息。

["支持的数据源文件格式"](#)

2025年2月2日

支持Amazon Nova基础模型

现在、GenAI支持Amazon Nova基础模型。支持Amazon Nova Micro、Amazon Nova Lite和Amazon Nova Pro。

["GenAI要求"](#)

数据源的文件类型筛选

现在、GenAI支持在添加数据源时选择要包括在数据源扫描中的特定文件类型。

"向知识库添加数据源"

数据源的文件修改日期筛选

现在、GenAI支持在添加数据源时按修改日期筛选要包含在数据源扫描中的文件。您可以为包含的文件选择修改日期范围。

"向知识库添加数据源"

支持图像文件和增强的PDF文件支持

现在、GenAI支持通过从图像和图形说明以及文档文本中获得洞察力来增强对知识库查询的响应、从而获得更丰富、质量更高的答案。现在、GenAI可以扫描PDF文件中的图像文件和图像(也称为多模式文件支持)。如果您选择扫描图像或PDF文件、则图像中的文本(包括PDF文档中嵌入的图像)将扫描到数据源中、扫描中的见解将包括在知识库查询的响应中。

"向知识库添加数据源"

混合搜索和重新搜索支持

现在、GenAI可以使用混合搜索并重新排列结果、从而显著提高搜索结果的相关性和准确性。混合搜索将基于关键字的传统搜索的优势与基于密集矢量的高级语法搜索技术相结合。标准关键字搜索结果通过近似匹配和语言细微差别得到增强、从而增强相关性。然后、GenAI使用Cohere Rerank和Amazon Rerank等高级重新排名模型进一步细化这些结果、并返回最相关的结果。此功能可用于新创建的知识库。

"了解NetApp Workload Factory for GenAI"

2025年1月5日

自定义快照名称

现在、您可以为临时快照提供快照名称。

"使用快照保护知识库"

自定义AI引擎实例名称

现在、您可以在部署期间为AI引擎实例提供一个自定义名称。

"部署GenAI基础架构"

重建损坏或缺失的GenAI基础架构

如果您的AI引擎实例损坏或以某种方式被删除，您可以让Workload Factory为您重建它。重建完成后，工作负载工厂会自动将您的知识库重新连接到基础设施，以便随时可以使用。

"故障排除"

2024年12月1日

从快照克隆一个信息库克隆

NetApp Workload Factory for GenAI 现在支持从快照克隆知识库。从而可以快速恢复知识库，并利用现有数据源创建新的知识库，有助于数据恢复和开发。

["克隆知识库"](#)

内部ONTAP集群发现和复制

发现并将本地ONTAP集群数据复制到 FSx for ONTAP文件系统，以便用于丰富 AI 知识库。所有本地发现和复制工作流程均可通过存储清单中新的 **On-Premises ONTAP** 菜单实现。

["发现内部 ONTAP 集群"](#)

2024年11月3日

使用数据防护屏蔽个人身份信息

生成式人工智能工作负载引入了由NetApp控制台分类提供支持的数据护栏功能。数据护栏功能可识别和屏蔽个人身份信息 (PII)，帮助您保持合规性并加强敏感组织数据的安全性。

["创建GenAI知识库"](#)

["了解NetApp控制台分类"](#)

2024年9月29日

为知识库卷提供快照和还原支持

现在、您可以通过创建知识库的时间点副本来保护生成性AI工作负载数据。这样、您就可以保护数据、防止意外丢失或测试对知识库设置所做的更改。您可以随时还原知识库卷的先前版本。

["为知识库卷创建快照"](#)

["还原知识库卷的快照"](#)

暂停计划的扫描

现在、您可以暂停计划的数据源扫描。默认情况下、生成性AI工作负载每天都会扫描每个数据源、以便将新数据导入到每个知识库中。如果您不希望(例如在测试或还原快照期间)导入最新更改、您可以随时暂停计划的扫描并恢复它们。

["管理知识库"](#)

现在、知识库可支持数据保护卷

现在、在选择知识库卷时、您可以选择属于NetApp SnapMirror复制关系的数据保护卷。这样、您就可以将知识库存储在已受SnapMirror复制保护的卷上。

["确定要集成到知识库中的数据源"](#)

2024年9月1日

其他分块策略

生成型AI工作负载现在支持对数据源执行多句分块和基于重叠的分块。

为每个知识库提供专用卷

现在、生成性AI工作负载会为每个新知识库创建一个专用的Amazon FSx for NetApp ONTAP卷、从而为每个知识库启用单独的快照策略、并提高对故障和数据中毒的防护能力。

2024 年 8 月 4 日

Amazon CloudWatch Logs集成

生成型AI工作负载现已与Amazon CloudWatch Logs集成、使您能够监控生成型AI工作负载日志文件。

聊天机器人应用程序示例

NetApp Workload Factory GenAI 示例应用程序使您能够通过基于 Web 的聊天机器人应用程序中直接与已发布的NetApp Workload Factory 知识库进行交互来测试其身份验证和检索。

2024年7月7日

GenAI 工作负载工厂的初始版本

初始版本支持开发一个知识库、该知识库可通过嵌入组织的数据进行自定义。您的用户可以通过聊天机器人应用程序访问知识库。此功能可确保准确、相关地回答组织特定的问题、从而提高所有用户的满意度和工作效率。

版权信息

版权所有 © 2025 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，本文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。