



利用 MongoDB 在应用程序中嵌入生成式人工智能和高级搜索

构建 AI 赋能的应用程序

2023 年 12 月

美国 866-237-8815 • 国际电话 +1-650-440-4474 • info@mongodb.com

2023 MongoDB, Inc. 保留所有权利

目录

简介	3
情境决定一切	3
向量和相似性搜索的兴起	4
向量搜索和大型语言模型 workflow	5
充满活力的 AI 生态系统的前景和现实	6
开发者数据平台：以智能方式构建智能应用程序	7
直接展示，无需多言。开发者数据平台上的生成式 AI 增强型应用程序	9
支持客户自助服务的聊天机器人和问答助手	9
高级电子商务搜索和推荐	12
富媒体（多模式）分析和生成	14
MongoDB 向量搜索的实际应用	14
入门	16

简介

从未有一项新技术的推出能如此迅速地引起企业、政府和消费者的关注。ChatGPT 于 2022 年 11 月一经推出，便展现了由大型语言模型 (LLM) 提供支持的生成式 AI 在解决大量新使用案例方面的潜力。这些使用案例以前是传统计算和分析 AI（现在有时称为“传统”或“经典” AI）无法想象的。

似乎只需要输入一些精心设计的提示，就可以自动完成各种各样的任务，从生成专业品质的文本、图像、音频、视频和编程代码，为客户提供更好的支持，到气候变化建模、发现新药物、设计新材料、预测金融市场走势，等等。

一夜之间，每家企业的董事会议程中都出现了这样一个问题 — “*我们如何使用生成式人工智能来颠覆市场，同时又避免被颠覆*”？

然而，技术领导者们很快认识到，除了潜在优势以外，生成式 AI 这项尚不成熟的技术也蕴含着风险。他们不能直接抛弃多年以来形成的运营最佳实践和系统知识，而是需要确保现有系统以及正在开发的新应用程序能够以安全、可靠和准确的方式利用生成式人工智能。

在本文中，我们将探讨 MongoDB 如何引导您走上实现这些目标的路径，同时利用自己的数据来增强生成式 AI 赋能的全新应用程序和体验。

情境决定一切

当每个人都可以访问 GenAI 模型时，您的“超能力”差异化来自于让这些模型访问您最重要的企业资产之一 — 您的数据。这些数据有些是组织专有的，有些是公开的，但比用于训练原始基础模型的数据更新。这些数据可以更好地反映当前的“实际情况”。

要为模型提供您自己的数据，这就需要采用一种新的架构模式，即检索增强生成 (RAG)。使用 RAG 可以为您的开发者提供一个强大的组合。开发者可以利用预先训练的通用生成式 AI 模型的令人难以置信的知识和推理能力，并为该模型提供准确且最新的企业相关数据。

结果是 GenAI 输出准确、最新、相关，并可利用您的所有数据，无论其结构如何。GenAI 驱动的应用程序可以更好地服务客户、提高员工工作效率并超越创新竞争对手。开发者可以解锁

所有这些结果，而无需求助于专门的数据科学团队来训练或微调模型 — 这是一个复杂、耗时且昂贵的过程。

使用您自己的数据源是让生成式人工智能为企业所用的重要一环。但仅靠这一点是远远不够的。正如本文稍后部分所述，开发者还需要考虑如何围绕基于数据的大型语言模型来部署应用程序，同时采取适当的安全控制措施，并确保达到用户期望的可扩展性和性能水平。

向量和相似性搜索的兴起

为了向 AI 模型提供我们自己的数据，我们需要首先将数据转换为向量嵌入。这些向量为我们的数据提供多维数字编码，以捕获其模式、关系和结构。向量嵌入为我们的数据赋予语义含义；通过计算向量之间的距离，我们的应用程序可以轻松理解不同数据对象之间的关系和相似性。这会将我们的数据开放给一系列全新的应用程序，具体将在下文中讨论。

任何数字格式和任何结构的数据，如文本、视频、音频、图像、代码、表格，都可以通过合适的向量嵌入模型处理后转化为向量。例如，OpenAI 的 `text-embedding-ada-002` 就是最常用的文本内容向量化模型之一。向量嵌入的美妙之处在于，非结构化数据对于计算机是完全不透明的，但现在可以通过嵌入来推断和表示其含义和结构。这意味着，我们可以开始像以往处理结构化业务数据那样来搜索和计算非结构化数据。我们每天创建的数据中有 80% 以上都是非结构化数据，从这一点就可以开始认识到向量搜索与生成式 AI 相结合所带来的巨大变革性潜力。

如下面的图 1 所示，一旦我们的数据被转换为向量嵌入，它就会在向量存储（如 [MongoDB Atlas Vector Search](#)）中持久化和编入索引。要检索类似的向量，使用近似最近邻 (ANN) 算法查询存储，以使用“分层可导航小世界” (HNSW) 等算法执行 K 最近邻 (KNN) 搜索。

这种向量查询能力让我们能够以处理前所未有的方式来处理数据，而这在以前只能依靠成本高昂的数据科学技能和基础设施。首先，我们可以将信息搜索和发现从关键字匹配扩展到情境感知语义搜索，后者能够从用户的搜索词中推断出含义和意图。其次，我们可以检索自己的数据（编码为向量），为生成式 AI 模型提供所需的情境，以便生成更可靠、更准确的输出。这些输出可以包括：

- 自然语言处理 (NLP)，用于支持聊天机器人、文本摘要问答和情感分析等任务。
- 计算机视觉和音频处理，用于支持图像分类和对象检测，以及语音识别和翻译。
- 内容生成，包括创建基于文本的文档和 SEO 优化的网页、计算机代码或将文本转换为图像或视频。

向量搜索和大型语言模型工作流

图 1 汇总了为大型语言模型启用“检索增强生成”的工作流。

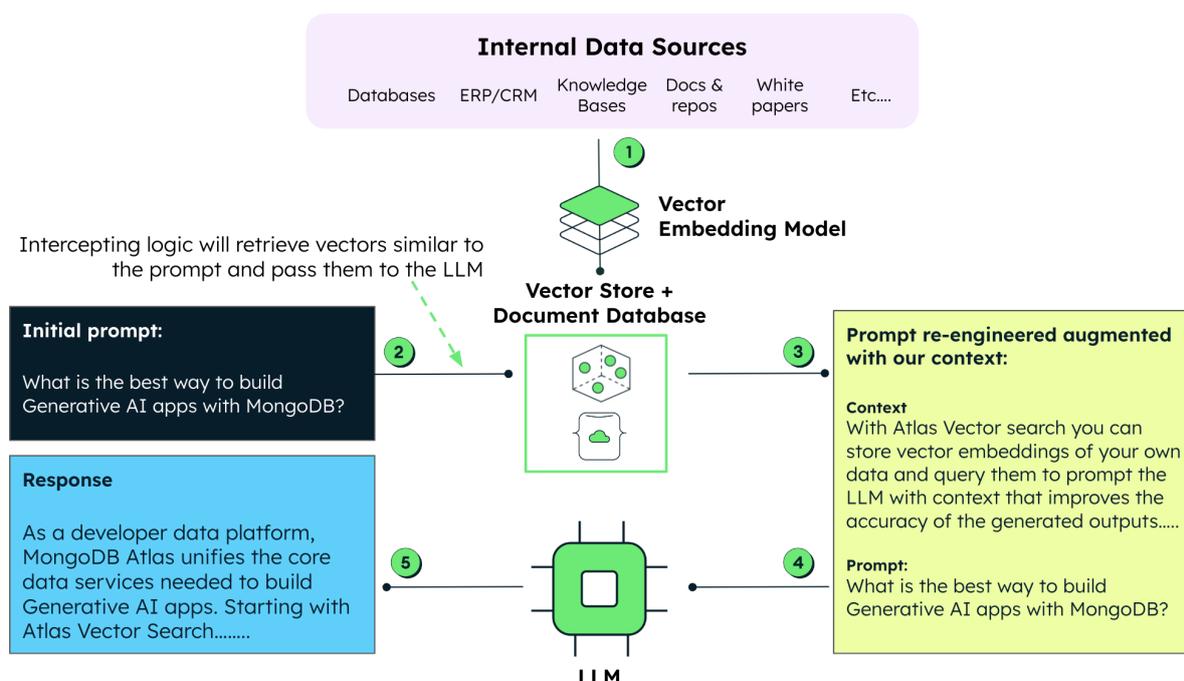


图 1: 将自定义数据与 LLM 动态组合，以生成可靠且相关的输出

我们的数据预先通过向量嵌入模型进行转换并存储在向量存储中。理想情况下，向量的元数据和原始“分段”数据与向量本身一起存储在一个灵活的文档数据库中，该文档数据库还存储常规应用程序数据。这让我们的应用程序能够以多种方式查询数据，提高相关性（例如对最近的数据给出更高的分数），并为大型语言模型提供长期记忆。输入给大型语言模型的提示会被从向量存储中检索相似向量的逻辑拦截，然后用来重新设计初始提示。新的提示会发送给大型语言模型，后者可以根据所提供的情境，使用更新的数据来生成更高质量和更准确的响应。

本文在稍后部分通过一些示例演示了上述工作流，并展示了如何将最终的功能应用于不同类别的应用程序。

充满活力的 AI 生态系统的前景和现实

向量存储是快速发展的 AI 赋能型技术生态系统的一部分，这是一个广阔的生态系统，涵盖嵌入创建、提示工程、大型语言模型、模型调优、编排、日志记录、基础设施自动化，等等。

这个生态系统中，各种富有吸引力和发展前景的项目和供应商都是企业的潜力合作对象，其中一些通过演示和原型展示了技术可能性。但企业决策者和开发者担心这些原型是否能轻松适应其特定业务需求。他们还担心一些较新的技术是否能够真正在任何操作环境中日复一日地以高水平的可靠性、可扩展性和安全性来保持生产负载正常运行。另一个问题就是如何集成组织的自主数据库，以将真实的业务数据输入到模型中。

AI 生态系统并不是孤立存在的。所有这些技术都需要嵌入到现实世界的应用中，才能真正为企业所用。例如，向量存储对于实现情境感知的生成式人工智能和语义搜索至关重要。但这些都只是整体应用程序的一部分，该应用程序还必须管理常规的非向量化业务数据。

这些数据可以是任何内容 — 客户记录、订单和库存、交易和事务、报价、地理空间坐标、产品详细信息和定价、时间序列测量和传感器读数、点击流和社交信源、文本描述等等。

所有这些数据都需要可供查询才能支持强大的应用程序功能。除了检索向量之间的近似最近邻以外，还应当能够完成各种常规操作，例如检索特定记录、处理大量数据更新以及运行支持分析处理的复杂聚合和转换。这种查询能力可以超越生成式 AI 的应用场景，为更多其他应用程序功能提供强有力的支持。而如果可以与模型的情境相关提示相结合，查询就可以提高生成式 AI 模型输出的准确性和相关性，从而发挥更加重要的作用。

除了处理应用程序数据和向量嵌入以外，我们还需要执行一些非功能性任务 — 满足正常运行时间、性能和可扩展性 SLA，集成新功能，以及数据保护、备份和审核。其中一些任务听起来很无聊。但如果这些任务发生失败，我们就会发现它们的重要性了.....

将这些技术结合在一起，为新的 AI 驱动型体验提供支持，并将它们融合到您的应用程序中，可能会造成单点产品和复杂性的蔓延，从而给您的团队带来巨大的开销。所有这些挑战加起来导致开发者体验支离破碎且效率低下、需要处理大量的操作和安全模型、大量的数据整理和集成工作以及大量的数据重复。所有这些都减慢了将 AI 驱动的新体验推向市场的速度，同时增加您的成本和风险。

基于 MongoDB Atlas 构建的开发者数据平台为您提供了一种更好的方法。

开发者数据平台：以智能方式构建智能应用程序

MongoDB 的开发者数据平台构建于 [MongoDB Atlas](#) 之上，统一了操作、分析和生成式人工智能数据服务，以简化构建智能应用程序的流程。无论您如何利用人工智能，从训练和服务自己的机器学习模型到在应用程序中嵌入最新的生成式人工智能，Atlas 都是您的技术堆栈的一个关键要素。从原型到生产，您可以利用 Atlas 自己的应用程序依托最新的操作数据，同时满足用户期望的可扩展性、安全性和性能水平。

MongoDB Atlas 的核心是其[灵活的文档数据模型](#)和开发者原生查询 API。两者共同助力您的开发者大幅加快创新速度，超越竞争对手并利用生成式人工智能创造的新市场机会。

文档是开发者处理数据的最佳方式，因为文档可以与代码中的对象建立映射关系，从而使其变得直观且易于推理。文档可以对任何结构的数据进行建模，从之前讨论的各种常规应用程序数据到由数千个维度组成的向量嵌入。所有这些结构都可以随时修改，以支持添加新的数据类型和应用程序功能。文档让您能够灵活地合理化和利用这些数据，而关系数据库的传统表格数据模型是无法做到这一点的。

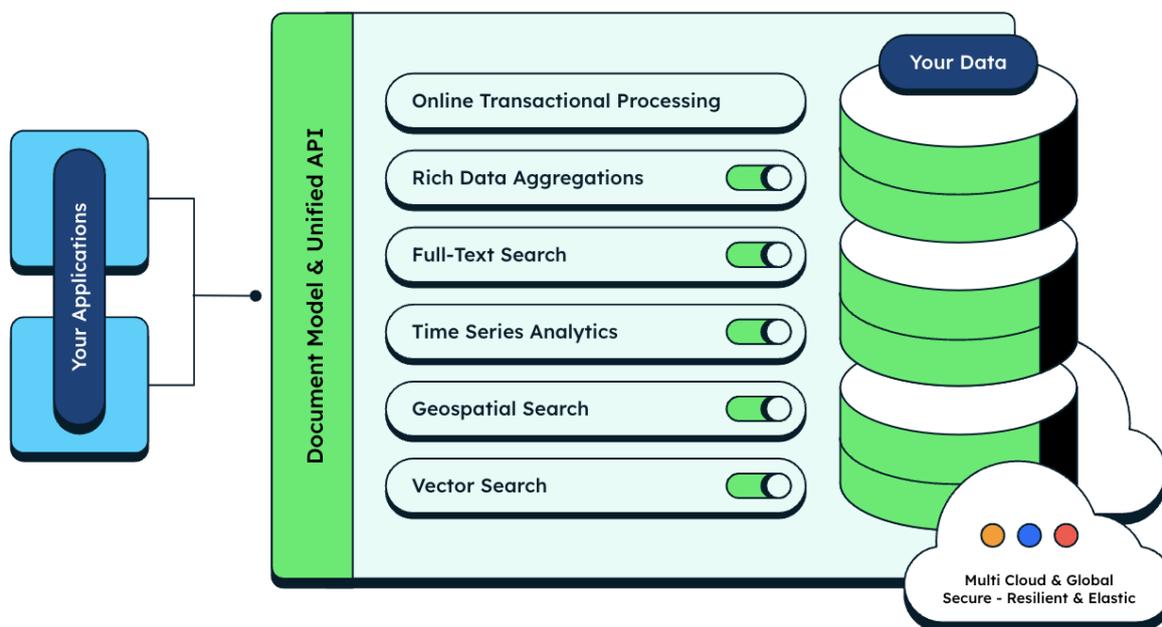


图 2: MongoDB Atlas 集成了将 AI 引入应用程序所需的数据服务

与文档模型相结合，[MongoDB Query API](#) 可为开发者提供了跨任何数据服务处理数据的统一且一致的方式。从简单的 CRUD 操作到关键字和向量相似性搜索，再到用于面向分析和流处理的复杂聚合管道，MongoDB Query API 让开发者能够灵活采用应用程序所需的任何方式来查询和计算数据。在生成式 AI 的背景下，这提供了高度灵活和强大的方法来定义基于矢量的查询的附加过滤器，例如：

- 结合元数据进行过滤：“为我查找与用户查询匹配的内容，但仅限于 X、Y、Z 年发布的内容”。
- 结合聚合：“查找与查询图像相似的所有图像，并按摄影师 ID 进行分组”。
- 与地理空间搜索相结合：“为我查找距离我所在位置 N 英里范围内与这张照片中房屋相似的房屋的房地产列表”。

任何其他数据库都无法在单一、统一的查询体验中提供如此广泛的查询功能。这让开发者能够更轻松地构建最终用户功能，复杂性大幅降低。开发者不再需要手动将来自多个数据库的查询结果拼接在一起，这是一个复杂、容易出错、成本高昂且速度缓慢的过程。同时，这还有助于将您的技术足迹控制在较低水平，并保持敏捷性。

“ MongoDB 已经在我们的系统中存储有关工件的元数据。随着 Atlas Vector Search 的推出，我们现在拥有了一个全面的向量元数据数据库，它经过了十多年的实战检验，可以解决我们密集的检索需求。无需部署新数据库，我们也不需要管理和学习新数据库。我们的向量和工件元数据可以放在一起存储。”

Pierce Lamb, [VISO TRUST](#) 数据和机器学习团队的高级软件工程师。

直接展示，无需多言。开发者数据平台上的生成式 AI 增强型应用程序

我们将重点通过三个常见使用案例来向开发者展示如何使用 MongoDB Atlas 构建 AI 赋能的应用程序：

- 支持客户自助服务的聊天机器人和问答助手。
- 高级电子商务搜索和用户推荐。
- 富媒体（多模态）分析和生成。

其中第一个示例都依赖于生成式人工智能和高级语义搜索来创建令人惊叹的用户体验，并让之前大多数组织都无法实现的功能成为可能。然而，要真正实现变革，这些 AI 增强功能需要成为整体应用程序的一部分，而该应用程序将为关键业务功能提供支持。

我们将依次介绍每个使用案例，展示其所依托的架构设计模式，以及 MongoDB Atlas 提供的相关功能。

支持客户自助服务的聊天机器人和问答助手

MongoDB 是许多客户支持应用程序的核心。这是因为 MongoDB 灵活的数据模型可支持轻松构建[单一的 360 度客户视图](#)。为此，MongoDB 可以从大多数组织中常见的无数孤立后端源系统中动态摄取多样且快速变化的客户数据。因此，基于 MongoDB 的单一整合式实时客户视图提供了一个理想的平台，让我们能够利用该平台来训练和实现聊天机器人及问答帮助功能，从而实现客户自助服务。

在图 2 所示的示例中，存储在 MongoDB 中的客户数据库作为 JSON 文件导出到嵌入式模型，该模型对数据进行分块（使用 LangChain 或 LlamaIndex 等工具）并从中创建向量嵌入。其他内部数据源（例如知识库和文档）也可以进行向量化以在应用程序中使用。然后将数据导入回 MongoDB 数据库。

为了确保通过最新的客户数据来持续更新向量，我们使用 [Atlas Triggers](#) 来观察单一视图中的任何数据变化。一旦在数据库中插入了新的客户记录或更新了现有记录，Atlas Triggers 就会调用嵌入模型的 API 来生成相应的向量并将其加载回 Atlas。

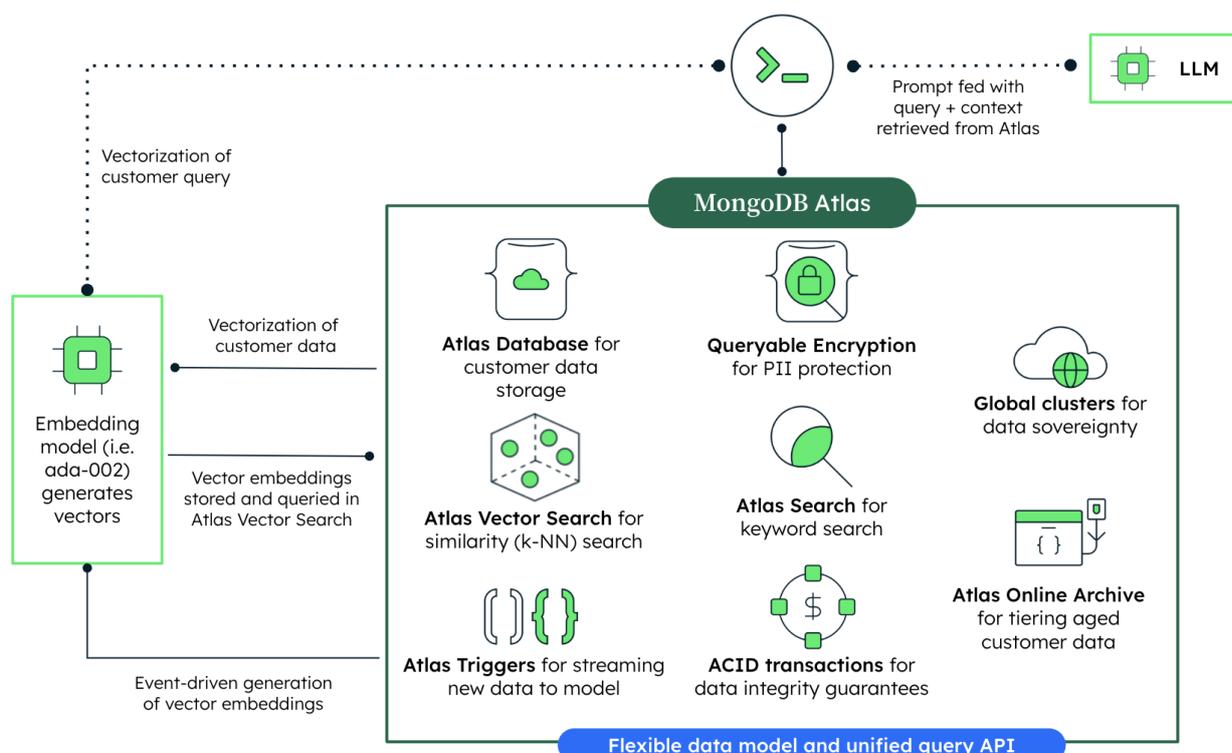


图 3: 由 MongoDB Atlas 提供支持的客户自助服务应用程序中内置的聊天机器人和 QA 生成式 AI 功能

借助 Atlas，开发人员可以利用 MongoDB 灵活的数据模型。开发者可以将源客户数据、元数据、数据段与向量嵌入存储在一起，所有数据均保持同步并以并排方式存储在单个存储层中，并通过单一查询 API 和驱动程序进行访问。

查询可以使用索引向量以及文档中常规字段的关键字索引来有效地过滤数据。这种集成意味着该应用程序可以支持更广泛的用户功能，同时降低开发者预算：

- [Atlas Vector Search](#) 通过对相应的索引嵌入数据执行相似性搜索来返回匹配文档。为了降低返回过时数据的风险，我们的查询可以使用向量的元数据（例如存储在 Atlas 数据库中的“创建日期”）来过滤掉较旧的内容。

- [Atlas Search](#) 根据来源和分段客户数据中的匹配关键字来返回结果。它使用模糊搜索等功能来纠正用户输入中的拼写错误，使用自动完成功能来提供建议的搜索词。它还使用索引交集来为针对客户数据的复杂临时查询提供高效支持。

对 Atlas 数据库、Vector Search 和 Atlas Search 的查询均使用相同的查询界面和驱动程序，大幅简化了开发者的工作流程。从 MongoDB Atlas 检索的数据作为上下文增强了提供给大型语言模型的提示，让大型语言模型能够对聊天和问题生成相关响应。上下文和提示以及用于回答复杂问题的任何相关推理步骤都会保存到 Atlas 中，为大型语言模型提供长期记忆并持续改进其输出。

客户数据是任何组织管理的最有价值的的数据之一。虽然生成式 AI 帮助我们在服务客户的方式进行创新，但保护他们的数据仍然至关重要。Atlas 提供了一系列功能来帮助我们做到这一点，使开发者能够专注于 AI 驱动的功能：

- 为数据存储、查询和分析、关键字搜索和向量搜索提供支持的融合基础架构。单一 API 和数据模型背后的这种统一大幅减少了开发者需要集成和构建的移动部件的数量。
- [Queryable Encryption](#) 是确保客户数据安全的行业首创。MongoDB 驱动程序会在客户端对敏感数据字段进行加密，数据库只会将其作为完全随机化的加密数据使用。即使对数据进行了加密，应用程序仍可对其运行表达式查询，而无需对数据库中的数据解密。需要注意的是，通常只有那些包含唯一标识个人的最敏感数据（如 SSN）的字段才会受到 Queryable Encryption 的保护。因此，可以对其余的明文字段进行搜索。
- Atlas 数据库中的[多文档 ACID 事务](#)保证了应用程序访问和修改客户数据时的完整性。
- 借助 [Atlas 全球集群](#)，客户数据可以锚定到其居住区域，从而符合现代数据主权法规。
- [Atlas Online Archive](#) 提供的完整的数据生命周期管理。该服务自动将活跃数据库中的过时客户数据分层到成本较低的云对象存储中，同时保持数据可供查询。对于在受监管行业中运行的应用程序中管理的客户数据，这一点非常重要。在这些行业中，客户数据必须保留多年，并确保可供访问。
- 通过备份和时间点恢复，保护客户数据免受损坏和勒索软件的侵害。

Atlas 在主要的超大规模云上为您提供全面管理，并以 99.995% 的正常运行时间 SLA 为后盾。

高级电子商务搜索和推荐

[电子商务产品目录](#)是 MongoDB 的一个常见应用场景：

- 各种不同的产品及其属性可以自然地映射到 MongoDB 灵活的文档数据模型中。
- Atlas 的分布式架构具有弹性横向扩展能力，让开发者能够动态调整数据库容量，以响应购物季和促销活动期间的应用程序需求。
- Atlas Search 的模糊搜索、自动完成、分面、突出显示和自定义评分等关键字匹配功能能让购物者能够快速浏览和导航产品目录，从而提高点击率 (CTR) 和购买转化率。

但是，关键字搜索需要匹配索引文本字段中的特定字词才能返回相关结果。如果未进行广泛而费力的同义词映射（例如，将单车映射到自行车或将运动鞋映射到训练鞋），当用户的搜索查询无法返回相关产品时，用户很快就会感到沮丧。这种挫败会导致销售损失和品牌声誉受损。

另一项挑战是向用户提供建议。开发者要么编写基于规则的复杂引擎，要么寻求专业且稀缺的数据科学资源。通常，数据首先必须从操作数据库提取、转换、加载 (ETL) 到离线数据仓库或数据湖。只有这样，传统的分析 AI 模型才能生成一系列建议，然后必须将其加载回操作数据库。这是一个复杂且开销很高的过程，所生成的推荐也是过时的，因为并不能反映用户最新的浏览行为或购买行为。

利用向量嵌入增强产品目录可以解决这些难题。

向量为我们目录中的产品提供语义含义，以便轻松理解产品之间的相似性和关系。这让商家能够以更低的工作量、复杂性和成本向用户展示相关产品。常见搜索词可以缓存在 MongoDB Atlas 中，从而更快地为用户提供相关结果。

通过将向量化扩展到客户数据（如前面的客户自助服务应用程序所示），我们可以结合产品和客户相似性搜索来优化建议，从而构建更复杂的推荐。

图 4 显示了高级搜索和推荐的整体设计模式。创建和维护向量嵌入遵循与上文所述客户自助服务应用程序中的聊天机器人和问答助手相同的工作流程。

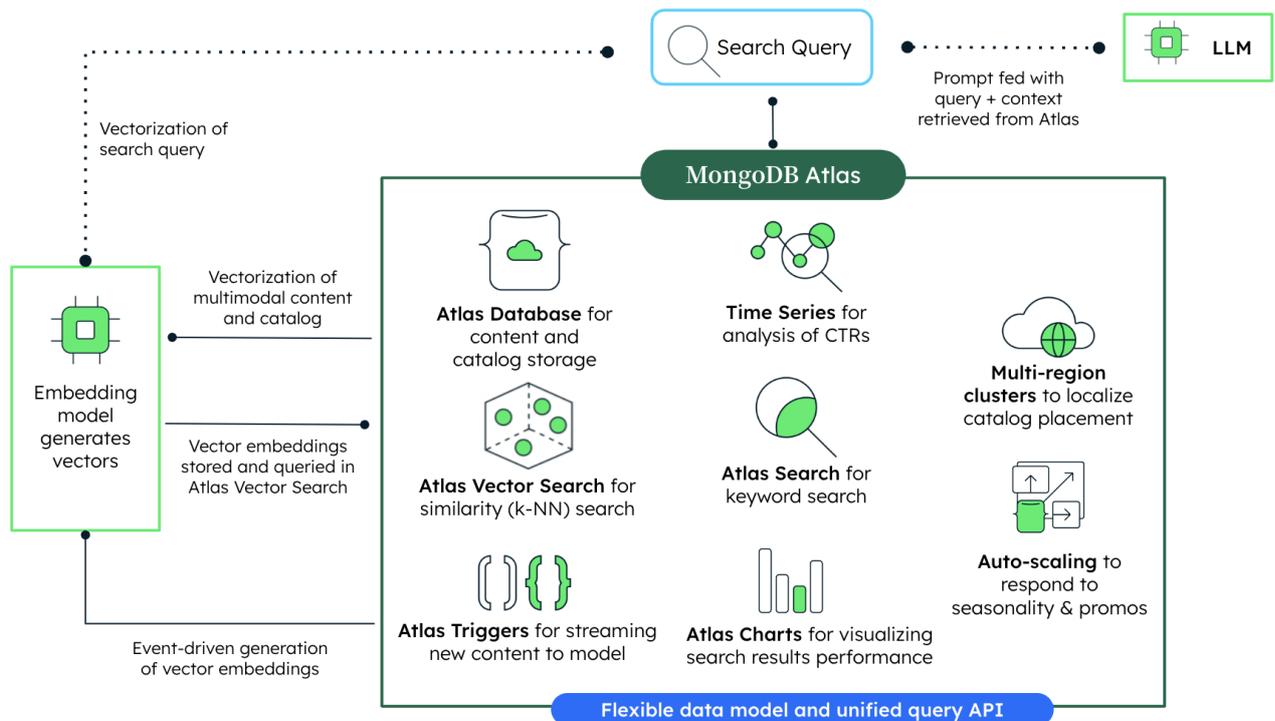


图 4：针对产品目录的高级语义搜索有助于提高销售转化率和新增销售

简单来说向量搜索如何显著改善产品搜索和推荐。整合 LLM 可以进一步增强这种体验。现在，客户可以提出实时问题并立即获得有关他们正在评估的产品的答案，从而有助于加快购买周期。

销售人员还可以使用大型语言模型来完成以前的各种人力密集型任务，这样他们就可以腾出时间开发更具创造性的方案来吸引客户。例如，大型语言模型可用于生成产品文案和 SEO 关键字的不同变体，然后可以对其进行 A/B 测试以量化哪种变体有助于实现更高的转化。大型语言模型可用于总结多条用户评论并推断情绪，帮助综合反馈意见，为产品路线图提供信息。

组织可以使用 Atlas 来管理整个电子商务生命周期。除了利用 AI 为搜索体验赋予更高的智能和预测性以外，企业主还可以从搜索结果中跟踪用户点击率和销售转化率。[时间序列集合](#)可以高效地从用户会话中摄取和存储高速和海量点击流，将这些数据用于分析以衡量搜索性能，包括使用 [Atlas Charts](#) 对结果进行实时可视化。借助这些见解，销售人员可以持续调整和优化产品数据与相关性评分，从而最大限度地提高电子商务网站的销售额。

富媒体（多模式）分析和生成

传统的关键字搜索可以很好地满足常规文本搜索的需要。然而，处理图像、语音和视频等富媒体（有时称为多模式）资产需要高度复杂的数据科学技术和技能。直到现在为止。

如前所述，任何数字内容都可以通过适当的向量嵌入模型进行向量化。[Hugging Face](#) 等 AI 中心以及超大规模云提供商的 AI 中心提供了各种针对不同内容模式进行了调优的模型。这些模型的嵌入可以存储在 Atlas Vector Search 中，从而为一系列新功能提供支持。如前所述，其中部分示例包括根据文本生成图像、转录视频以进行语音识别和情感分析以及图像分类和检测物体。来自不同媒体的向量可以组合在一起，例如通过比较文本和图像嵌入，以检查给定的句子是否准确地描述了图像。

这种多模态功能可用于广泛的应用场景。例如，丰富产品目录（如上所述）以及通过分析图像和视频来增强发现能力。这些功能可用于简化设计、制造和发布流程，或在安全和监控或增强现实 (AR) 等领域创建全新的应用程序类别。

针对上述高级电子商务搜索和推荐描述的架构设计模式和 MongoDB Atlas 功能同样适用于多模态内容生成。

MongoDB 向量搜索的实际应用

MongoDB 已经广泛应用于传统的 AI 应用场景。德国大陆集团 (Continental) 在其 [“零碳愿景”自动驾驶计划](#) 中选择了 MongoDB 来支持为其特征工程平台。[Bosch](#) 和 [Telefonica](#) 都在其 AI 赋能的物联网平台中采用了 MongoDB。[Kronos](#) 使用通过 MongoDB 数据配置和构建的机器学习模型，每天处理数十亿美元的加密货币交易。[Iguazio 使用 MongoDB](#) 作为其数据科学和 MLOps 平台的持久层，而 H2O.ai 和 Featureform 则在各自的平台中使用 MongoDB 为存储特征数据。

在此基础上，MongoDB Atlas 目前已广泛应用于多个应用程序，并正在助力这些应用程序突破生成式 AI 的可能性界限。请查看我们的[案例研究页面](#)，详细了解 MongoDB Atlas 所支持广泛的应用场景。其中一部分具体示例包括：

- [Ada](#): 通过 AI 驱动的自动化和对话式 AI，助力 Meta、ATT 和 Verizon 等企业改善客户支持。
- [ExTrac](#): 通过分析实时数据流，识别和分类新出现的物理和数字风险。
- [Eni](#): 解锁地质数据并将其用于改善决策，推动企业加速实现净零排放。
- [Inovaare](#): 对整个医疗生命周期的数据进行持续监控、提取和分类，以满足合规报告、审计和风险评估需求。
- [Source Digital](#): 视频检测平台从 PostgreSQL 迁移到 MongoDB Atlas 后，成本降低为原来的七分之一。
- [Catylex](#): 自动提取、分类和分析合同条款，以确定权利、义务和风险。
- [Robust Intelligence](#): 通过 AI 防火墙产品实时验证输入和输出，为生产环境中的大型语言模型 (LLM) 提供安全保障。
- [Potion](#): 使用自定义视觉和音频模型重新生成视频和音频流。

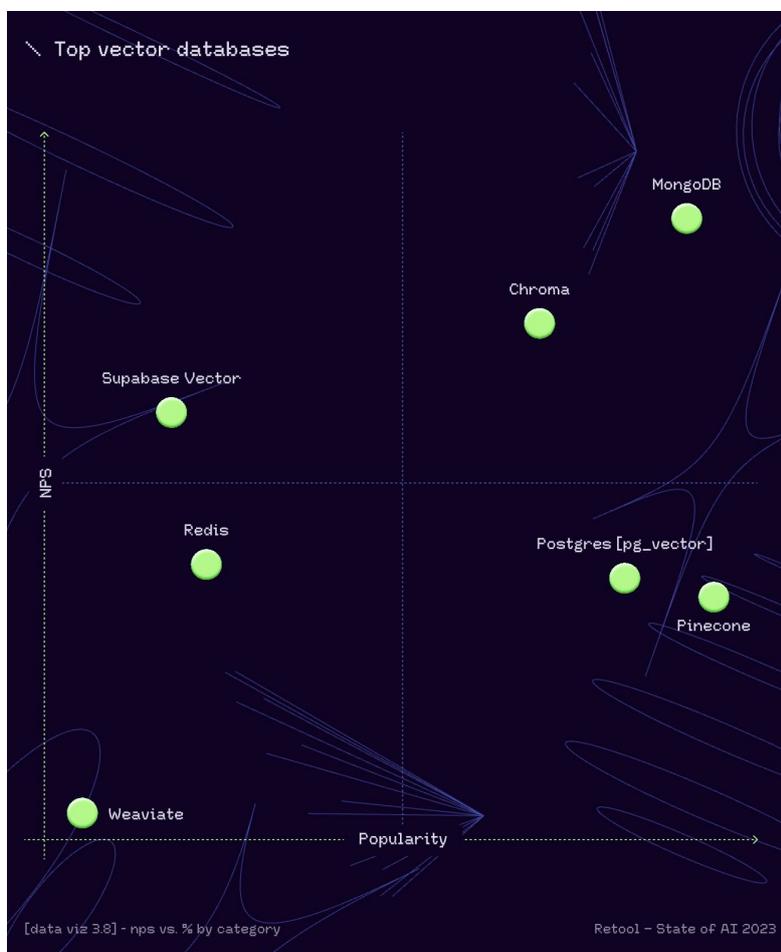


图 5: Retool AI 现状调查 — 行业领先的向量数据库

软件工具供应商 Retool 的 [AI 现状调查](#) 反映了 MongoDB 在 AI 开发者中的受欢迎程度，该调查表明 MongoDB Atlas Vector Search：

1. 在所有受调查的向量数据库中，净推荐值 (NPS) 最高。
2. 在发布后的短短几个月内就跃升为第二大最广泛使用的向量数据库，领先于许多已经存在多年的其他解决方案。

“Atlas Vector Search 稳健可靠、性价比高，而且速度极快！”

[Kovai 首席执行官 Saravana Kumar](#) 谈及其公司的 AI 助手开发。

入门

无论您是在初创公司还是企业中打造下一个重大项目，MongoDB Atlas 都能助力您：

- 加速构建基于真实运营数据的生成式 AI 密集型应用程序。
- 利用单一平台简化您的技术堆栈，该平台可允许您的应用程序将操作数据和向量嵌入存储在单一位置，使用 serverless 功能对源数据的变化做出响应，并跨多种数据模态进行搜索，从而提高应用程序生成的响应的相关性和准确性。
- 借助文档模型的灵活性，轻松改进您的生成式 AI 密集型应用程序，同时保持简洁、优雅的开发体验。
- 无缝集成领先的 AI 服务和系统，例如超大规模服务提供商以及开源大型语言模型和框架，以在动态市场中保持竞争力。
- 在高性能、高度可扩展的操作数据库上构建生成式 AI 密集型应用程序，该数据库已经在各种 AI 使用案例中经过十年的验证。

您可以访问我们的 [AI/ML 资源中心](#)，详细了解如何使用 MongoDB 构建 AI 驱动的应用程序。

要开始使用 MongoDB Atlas，开发者只需在 [MongoDB Atlas](#) 上注册一个帐户。然后，开发者就可以使用 Atlas 数据库、Atlas Vector Search 和 Atlas Search 创建免费的 MongoDB 实例，加载自己的数据或我们的样本数据集，并探索平台内的各种可能性。

免责声明

此处所述有关我们产品的任何特性或功能的开发、发布和日程安排完全由我们自行决定。此信息仅用于概述我们产品的总体发展方向，不应作为制定购买决策的依据，也并非提供任何材料、代码或功能的承诺、诺言或法律义务。

美国 866-237-8815 • 国际 +1-650-440-4474 • info@mongodb.com。

© 2023 MongoDB, Inc. - MongoDB 和 MongoDB 绿叶徽标均为 MongoDB, Inc. 的注册商标。