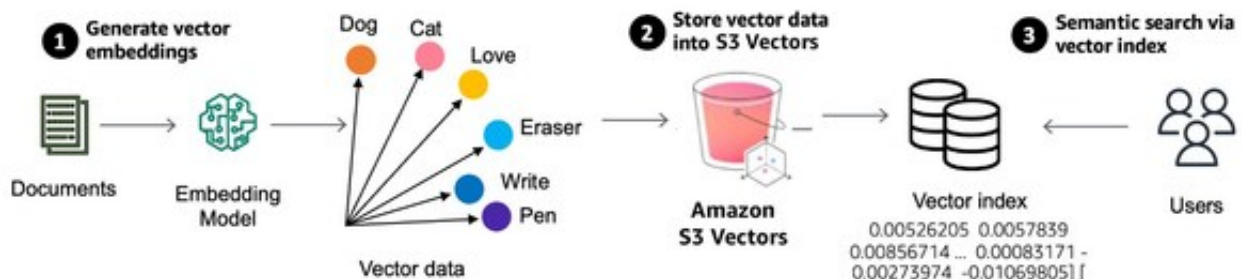


亚马逊云科技推出首个支持规模化原生向量处理的云存储服务

北京2025年7月28日 /美通社/ -- 亚马逊云科技日前宣布，推出Amazon S3 Vectors预览版，专为持久存储向量数据而打造，能够将向量数据的上传、存储和查询总成本降低多达90%。Amazon S3 Vectors是首款具备原生存储大规模向量数据集能力的云存储服务，可提供亚秒级查询性能，让企业能够以经济实惠的方式，大规模存储适用于AI的数据。该服务现可与Amazon Bedrock、Amazon OpenSearch Service和Amazon SageMaker Unified Studio服务集成，帮助企业以更高性价比构建语义搜索、检索增强生成（RAG）和AI Agent等生成式AI应用。



作为一种新兴技术，向量搜索在生成式AI应用中，通过利用距离或相似度度量标准，比较数据的向量表示形式，来查找与给定数据相似的数据点。向量则是利用嵌入模型生成的无结构化数据的数值表示形式，用户可以使用嵌入模型为文档中的各个字段生成向量，并将这些向量存储到Amazon S3 Vectors中，以便进行语义搜索。



Amazon S3 Vectors引入了向量存储桶，这是一种配备了一组专用API的新型存储桶，用户无需配置任何基础设施即可存储、访问和查询向量数据。创建Amazon S3 Vectors时，用户可在向量索引中对向量数据进行组织，从而能够轻松地对数据集运行相似性搜索查询。每个向量存储桶最多可拥有10,000个向量索引，且每个向量索引可存储数千万个向量。

创建向量索引后，在向该索引添加向量数据时，用户还可为每个向量附加上键值对形式的元数据，以便后续根据日期、类别、用户偏好等一系列条件，对查询结果进行筛选。随着时间推移，当用户不断对向量进行写入、更新以及删除等操作时，即便数据集规模持续扩大、内容不断演变，Amazon S3 Vectors也会自动优化处理向量数据，从而确保向量存储达到最佳性价比。

Amazon S3 Vectors可与Amazon Bedrock知识库及Amazon SageMaker Unified Studio集成，可用于构建极高性价比的检索增强生成（RAG）应用。Amazon S3 Vectors还通过与Amazon OpenSearch Service集成，可将查询频率较低的向量存储在Amazon S3 Vectors中，实现存储成本降低。并且随着查询需求增加，又能迅速将这些向量迁移至Amazon OpenSearch Service中；或者当需要支持实时、低延迟的搜索操作时，也能通过该功能轻松实现。

借助Amazon S3 Vectors，企业能以经济高效的方式，将代表海量无结构化数据（如图像、视频、文档和音频文件）的向量嵌入存储起来，从而使得可扩展的生成式AI应用成为可能，包括语义搜索、相似性搜索、RAG以及构建Agent记忆等应用。此外，企业还可以开发各类应用，轻松应对多种行业的应用场景需求，如个性化推荐、自动化内容分析和智能文档处理等，并且无需承担管理向量数据库的操作复杂性与管理高成本。

目前，Amazon S3 Vectors，及其与Amazon Bedrock、Amazon OpenSearch Service和Amazon SageMaker的集成功能，现已在美国东部（弗吉尼亚北部）、美国东部（俄亥俄）、美国西部（俄勒冈）、欧洲（法兰克福）和亚太（悉尼）区域提供预览版。

用户可在Amazon S3控制台中，立即开始试用Amazon S3 Vectors功能。

原文地址：<http://www.china-nengyuan.com/news/230783.html>