

解决方案简介

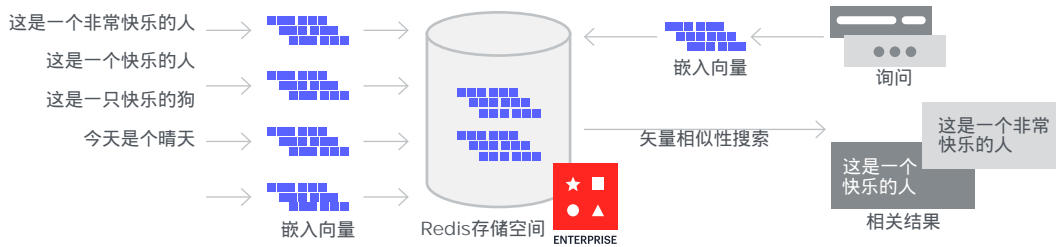
Redis Enterprise 向量相似性搜索

从传统搜索到向量相似性



向量相似性搜索

用户期望在他们遇到的每一个应用程序和网站中都有搜索功能。然而，超过80%的商业数据是非结构化的，以文本、图像、音频、视频或其他格式存储。AI驱动搜索功能被集成到Redis Enterprise中，以实现向量相似性搜索。



1 将原始数据转换为嵌入向量

嵌入向量是无结构数据的数值表示形式，能够捕捉语义信息，通常由人工智能模型创建。

图像、文本、视频甚至是一个音频，在 Huggingface Model Hub 的帮助下，每种内容都可以转换成各自的嵌入向量。

cohere, OpenAI, etc.

2 存储在 Redis 中

3 Redis 支持两种向量索引方法

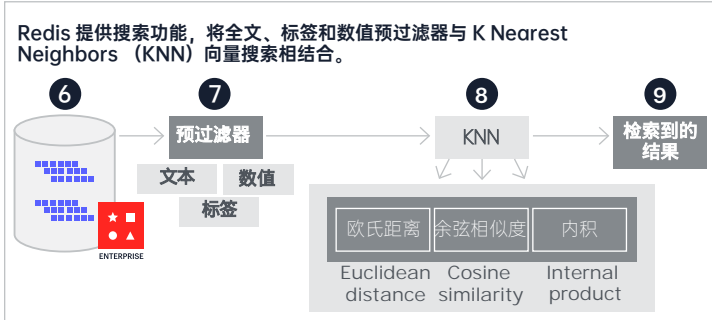
FLAT
一种 (brute force) 方法，它遍历所有可能的向量进行搜索

HNSW
一种近似搜索方法，能够以更快的速度获得结果，但精确度较低

4 两种方法都具有相同的必需参数。

类型	浮点数 (32位)
向量维度	128
距离度量	余弦相似度

5 索引只需要创建一次。当新的哈希数据被存储在 Redis 中时，它们会自动重新建立索引。



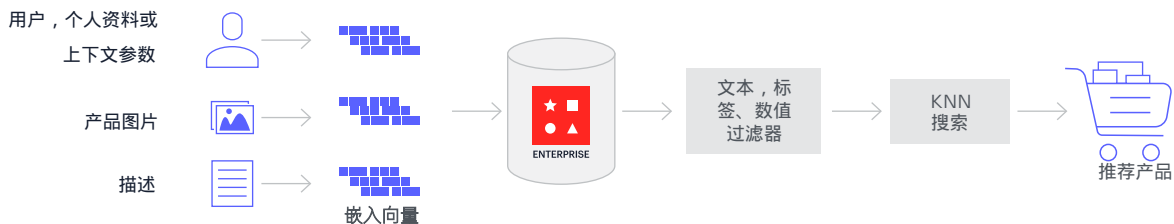
实现无可匹敌的性能和规模

- 每秒查询次数 (QPS) 增加了12倍
- 延迟降低至原来的1/13
- 99.999% 的高可用

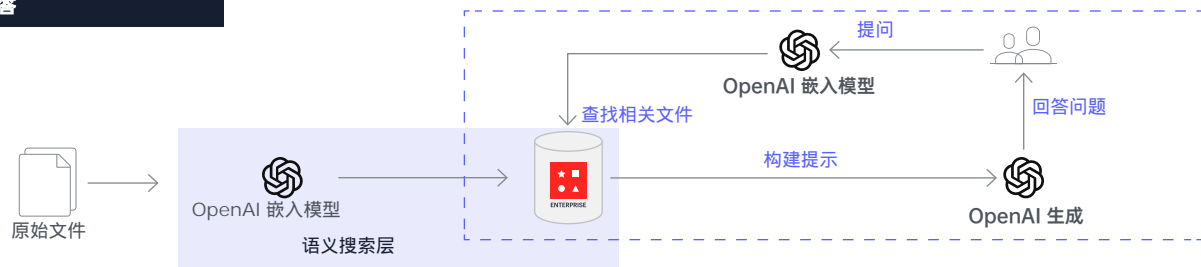
应用案例

- 文档检索
- 问答
- 推荐系统
- 视觉搜索

应用案例：推荐系统



应用案例：问答





虹科电子科技有限公司
T(+86)400-999-3848
M(+86)155 2866 3362
www.hongcloudtech.com
hongcloudtech@hkaco.com

广州市黄埔区神舟路18号润慧科技园C栋6层
各分部：广州 | 成都 | 上海 | 苏州 | 西安 | 北京 | 台湾 | 香港 | 美国硅谷



联系我们



行业交流群



hongcloudtech.com



获取更多资料