向量数据库

常见问题

腾讯云

版权所有：腾讯云计算（北京）有限责任公司

【版权声明】

©2013-2025 腾讯云版权所有

本文档（含所有文字、数据、图片等内容）完整的著作权归腾讯云计算（北京）有限责任公司单独所有，未经腾讯云事先明确书面许可，任何主体不得以任何形式复制、修改、使用、抄袭、传播本文档全部或部分内容。前述行为构成对腾讯云著作权的侵犯，腾讯云将依法采取措施追究法律责任。

【商标声明】

腾讯云

及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。未经腾讯云及有关权利人书面许可，任何主体不得以任何方式对前述商标进行使用、复制、修改、传播、抄录等行为，否则将构成对腾讯云及有关权利人商标权的侵犯，腾讯云将依法采取措施追究法律责任。

【服务声明】

本文档意在向您介绍腾讯云全部或部分产品、服务的当时的相关概况，部分产品、服务的内容可能不时有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或默示的承诺或保证。

【联系我们】

我们致力于为您提供个性化的售前购买咨询服务，及相应的技术售后服务，任何问题请联系 4009100100或95716。

文档目录

常见问题
产品功能常见咨询
实例计费相关咨询
AI 套件服务相关咨询
API 与 SDK 常见问题

版权所有：腾讯云计算（北京）有限责任公司

常见问题
产品功能常见咨询

最近更新时间：2025-04-27 16:13:52

如何新建向量数据库?

1. 选择适合自身业务需求的数据库实例规格，具体信息，请参见 规格选型 。
2. 掌握数据库实例计费的方式、计费项目表、计费公式，以及相关续费、退费、欠费说明等。具体信息，请参见 计费概述 。
3. 按照所选择的规格、所选择的计费方式，购买业务所需的数据库实例。具体操作，请参见 购买实例 。

图像、文本、音视频等这些非结构化数据如何转换为向量数据来存储?

- AI 套件是腾讯云向量数据库（Tencent Cloud VectorDB）提供的一站式文档检索解决方案，包含自动化文档解析、信息补充、向量化、内容检索等能力。用户仅需上传原始文档，数分钟内即可快速构建专属知识库，大幅提高知识接入效率。具体信息，请参见 AI 套件 。
- Embedding 功能是腾讯云向量数据库（Tencent Cloud VectorDB）提供将非结构化数据转换为向量数据的能力，目前已支持文本 Embedding 模型，能够覆盖多种主流语言的向量转换，包括但不限于中文、英文。具体信息，请参见 Embedding 。
- 图像、音频当前暂不支持，后续将逐步推出，请关注 产品动态 。

连接向量数据库失败，如何处理?

- 向量数据库支持内网访问，如果连接失败，建议您从以下方面进行排查：
  - 请检查云服务器 CVM 和向量数据库实例是否**内网互通**。
    - CVM 与 向量数据库是否属于同一 VPC。
    - 安全组是否配置正确。使用腾讯云 CVM 连接腾讯云向量数据库（Tencent Cloud VectorDB），在腾讯云 CVM 安全组中需配置**出站规则**，把腾讯云向量数据库的 IP 及端口添加到出站规则中。在腾讯云向量数据库安全组中配置**入站规则**，把 CVM 的 IP 地址及向量数据库的端口添加到入站规则中，才能连接成功。
  - 请求中的 API Key 是否获取正确。
- 向量数据库支持**外网访问**，如果连接失败，建议您从以下方面排查：
  - 请检查外网访问的白名单列表是否配置正确。具体操作，请参见 开启外网访问 。
  - 请确认 API Key 是否配置正确。

如何选择适合自己应用的向量数据库?

版权所有：腾讯云计算（北京）有限责任公司

购买实例之前，请了解向量数据库所支持的实例类型、节点类型与节点规格，选择与业务场景高度匹配的数据库实例。具体如何评估并选择，请参见 规格选型 。

如何在向量数据库中写入数据?

1. 掌握数据库设计的逻辑架构，使用向量数据库的 SDK 与 API 接口写入数据都是遵循逻辑架构而设计。具体信息请参见 设计架构 。
2. 掌握向量数据库支持的索引类型、与相似度计算方式，以便更顺利地应用 SDK 与 API 接口。具体信息请参见 索引与计算 。
3. 掌握向量数据库支持的检索方法，以便在不同的场景选择合适的检索方法。具体信息请参见 检索方法 。
4. 掌握控制台管理数据库实例的操作指引，以便快速高效管理实例的生命周期。具体操作请参见 管理实例 。
5. 向量数据库提供 Python SDK、Java SDK、Go SDK 以及 HTTP API ，辅助您快速访问向量数据库。第一次体验数据库的写入检索能力，请参见 第一次相似性检索 。
6. SDK 安装、接口应用示例、错误码返回信息，请参见 SDK 参考 。

如何使用向量数据库进行相似度搜索?

- 相似度检索是基于多维向量进行 相似度计算 的检索方式。相似度计算通过计算查询向量与数据库存储的向量之间的相似度，找到与查询的多维向量最相似的文档。当前支持的检索方式，请参见 检索方法 。
- 第一次体验数据库的相似性检索能力，请参见 第一次相似性检索 。
- Python SDK、Java SDK、Go SDK 以及 HTTP API 相似性检索的接口，请参见 SDK 参考 。

与传统的数据库相比，向量数据库有哪些特点?

向量数据库更适用于 AI 运算、检索场景，数据接入效率是传统方案的10倍。相较传统数据库具体有以下几个特点：

- 数据规模不同，能够高效处理大规模数据

  对于传统数据库而言，1亿条数据已经是很大的业务流量。然而在向量数据库面向的场景中，单索引数据量可能达到千万级、甚至亿级别，单条向量数据的维度也会达到上千维。
- 查询方式不同，支持近似查找

  传统数据库的查找方式都属于精确查找，而向量数据库通常是近似查找，即返回和输入内容最相近的 TOP K 条数据。
- 场景不同

  向量数据库更适用于 AI 运算、检索场景，能够高效支持大模型的接入和运算。

腾讯云向量数据库怎么删除实例资源?

1. 在向量数据库控制台，操作退还实例资源。具体信息，请参见 自助退还实例 。
2. 相关退还计费规则，请参见 退费说明 。查看退费进度与明细，请参见 查看费用账单 。

插入的向量数据为何有精度损失?

向量数据库支持存储的向量数值类型为单精度浮点型（Single-precision floating-point），又称 FP32 或 float32。单精度浮点型具有32位（即4字节）的宽度。这32位被分为三个部分：

- 符号位（Sign bit）：1位，用于表示数值的正负。
- 指数位（Exponent bits）：8位，用于表示数值的范围。
- 尾数位（Mantissa or Fraction bits）：23位，用于表示数值的精度。

若传入的向量数据精度超出上述表达范围，将自动近似到单精度表达范围内的最近值，也就发生了“精度损失”。

版权所有：腾讯云计算（北京）有限责任公司

实例计费相关咨询

最近更新时间：2024-12-26 15:15:02

向量数据库计费模式是否支持转换?

计费模式支持包年包月计费与按量计费。若购买实例时选择按量计费方式，购买后支持变更为包年包月计费。更多信息，请参见 按量计费转包年包月 。

包年包月计费即将到期，如何续费?

正式购买的包年包月实例即将到期的实例，您需要及时进行续费，避免资源到期无法使用而影响业务。续费操作，请参见 续费（包年包月） 。

隔离在回收站的实例，如何恢复?

主动退还实例、包年包月实例到期或者按量计费实例账户余额不足时，实例均被移入在回收站隔离，在保留时长内账户余额充足的情况下，均可以进行恢复。回收站实例的保留时长，不同计费模式有所不同。具体信息以及恢复操作，请参见 恢复已隔离实例 。

购买的实例不再需要，退还之后，如何退款?

- 包年包月预付费实例到期之前，如不再需要实例资源需退还实例时，可在控制台直接发起退还流程，结算退款金额。具体操作，请参见 自助退还实例 。
  - 5天无理由自助退还的金额将退还至腾讯云账户。
  - 普通自助退还的金额将按购买支付使用的现金和赠送金支付比例退还至您的腾讯云账户。
  - 推广奖励渠道订单退款将收取订单实际现金支付金额的25%作为退款手续费。 推广奖励渠道订单暂不支持自助退款，请通过 在线咨询 发起退款申请。
  - 自助退还计算规则、退费示例、退费进度及明细，请参见 退费说明 。
- 按量计费后付费实例，腾讯云账户未欠费时，如不再需要实例资源退还实例时，也可在控制台直接销毁实例。系统将按量结算需支付的费用，不涉及退款。

包年包月计费实例，到期未及时续费导致欠费影响业务，如何处理?

包年包月计费实例到期后8天开始，此云数据库不可再访问，将被隔离回收至**回收站**中。回收站保留时间最长为**七天**，这7天内，登录 向量数据库控制台 ，可在控制台**回收站**进行续费操作恢复实例。具体操作，请参见 恢复已隔离实例 。而若已经超过 7 天，资源已被清除无法再次恢复。

> **说明：**

> 登录 向量数据库控制台 在实例列表中，可看到实例状态以及实例到期时间。具体信息，请参见 查看实例信息 。到期后七天内云服务还可以继续使用，实例状态为**已过期**，系统将发送云服务到期提醒，还请您及时进行续费。

按量计费实例，腾讯云账户欠费，影响正常业务，如何处理?

按量计费实例在账户费用小于 0 之后，24小时内，云数据库可继续使用且继续扣费。24小时后，云数据库实例将**自动关机且停止扣费**，实例将会从实例列表移除并展示在**回收站**。

- 隔离7天内，账户余额 充值 为正后，按量计费实例将会被变为可恢复状态，此时您可以登录 向量数据库控制台 的回收站，对实例执行恢复操作。具体操作，请参见 恢复已隔离实例 。
- 隔离7天后，若您的账户余额仍为小于0的状态，云数据库将被下线，所有数据均会被清理，且不可找回。向量云数据库下线时，系统将通过邮件及短信等方式通知到腾讯云账户的创建者以及全局资源协作者、财务协作者。

如何查看向量数据库实例的费用明细?

您可以通过腾讯云控制台的**费用中心**，查看您的账户可用余额、收支明细、为您的账户进行充值汇款等。具体操作，请参见 查看费用账单 。更多信息，请参见 计费相关 介绍。

实例不再需要，已操作退还，退款金额未到账，如何查看进度?

1. 登录 向量数据库控制台 。
2. 在页面右上角，选择**费用** > **我的订单**，进入**费用中心**的**订单管理**页面。
3. 在**预付费订单**页签可随时跟踪包年包月退款到账进度。

购买时，开启了自动续费，购买之后，如何取消自动续费?

包年包月计费的实例，购买时，开启自动续费。购买之后，支持取消自动续费。具体操作，请参见 续费（包年包月）。

版权所有：腾讯云计算（北京）有限责任公司

AI 套件服务相关咨询

最近更新时间：2024-03-29 14:46:11

已购买的实例，为什么不能使用 create_ai_database 接口?

AI 套件是腾讯云向量数据库（Tencent Cloud VectorDB）提供的一站式文档检索解决方案，包含自动化文档解析、信息补充、向量化、内容检索等能力。

1. AI 套件当前支持地域包含：北京、上海、广州、新加坡。其余地域暂不支持。
2. 如果上传文件为 Markdown 类型，Markdown 类型文件最大限制为 1MB，请确认文件大小。
3. 当前实例为发布 AI 套件功能之前购买的实例，不支持 AI 套件功能，请 提交工单 申请升级实例版本。

已购买的实例，为什么不支持上传 PDF、Word、PPT类型?

AI 套件发布第一个版本仅支持 Markdown 类型，第二个迭代已支持 PDF、Word、PPT 类型文档。具体发布时间段，请参见 产品动态 。

- AI 套件当前支持地域包含：北京、上海、广州、新加坡，其余地域暂不支持。
- 确认上传文件的大小：PDF、Word、PPT 最大限制为 10MB。
- 确认实例的创建时间：如果为 2024-02-22 之前创建的实例，请 提交工单 申请升级实例版本，才能支持上传 PDF、Word、PPT。

如何获取一个 Collection 中所有的 Document 信息?

使用 query 接口，配置 offset 与 limit 参数，进行范围查询，获取整个数据库的所有表 ID 信息，再通过 get 接口逐一配置文档 ID 信息，逐一获取文档数据。

版权所有：腾讯云计算（北京）有限责任公司

API 与 SDK 常见问题

最近更新时间：2024-04-01 14:33:51

SDK 支持哪一些开发语言?

腾讯云向量数据库（Tencent Cloud VectorDB）的 SDK 是将 HTTP API 封装成易于使用的 Python 函数或类。开发者可以通过 Python SDK 更加方便地操作数据库。当前仅支持 Python、Java、Go 语言类型，其他语言类型暂不支持。后续建工会根据客户诉求逐一排期，请关注 产品动态 。

Python SDK 接口请求失败，如何处理?

Python SDK 使用异常调试技巧：请在代码执行前执行如下信息，打开 Debug 模式，获取运行日志，通过日志提示信息，结合 错误码 信息，分析原因。若还无法解决，请 提交工单 咨询腾讯云工程师。

```
import tcvectordb
tcvectordb.debug.DebugEnable = True
```

更新与插入文档数据差异是什么?

- 插入数据：插入数据时若指定 ID 已存在，已存在的数据不会被覆盖，本次插入数据操作提示无效。
- 更新数据：更新数据将会覆盖已存在数据，若指定 ID 不存在，则本次更新数据操作无效。

相似性检索数据时，是否支持批量操作?

- 相似性检索时，支持一次配置多个主键 id（Document ID）或多组向量数值，进行相似性数据，数组元素数量最大为20。
- 配置 Embedding 参数，支持输入原始文本检索与输入文本相似的文档，支持一次配置多条文本信息，进行检索，数据元素最大为20。
- 应用 AI 套件进行相似性检索，仅支持输入一条文本信息进行检索。

版权所有：腾讯云计算（北京）有限责任公司