

向量数据库 控制台指南



腾讯云

【 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 】



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

控制台指南

管理实例

查看实例信息

自助退还实例

恢复已隔离实例

立即下线

标签管理

配置变更

升配节点规格

增加节点数量

维护时间窗

监控实例

查看监控指标

配置监控告警

管理密钥

管理安全组

管理 Embedding 功能

登录 DMC

切换网络

开启外网访问

备份与回档

备份概述

查看备份记录

自动备份

手动备份

克隆实例

控制台指南

管理实例

查看实例信息

最近更新时间：2025-03-20 16:56:52

操作场景

购买腾讯云向量数据库（Tencent Cloud VectorDB）之后，您可以在控制台直观、快速查看实例的详细信息，包括：实例的运行状态、资源使用情况、网络状态等信息，并对实例进行高效运维管理。

前提条件

- 已成功 [新建数据库实例](#)。
- 实例未被销毁。

查看实例列表

1. 使用腾讯云账号登录 [向量数据库控制台](#)。
2. 在右侧页面上方，**向量数据库 - 实例列表**右边的下拉框选择地域。
3. 可看到当前账号下所选地域下的所有实例，在右上角的搜索框，可输入实例名称搜索目标实例。
4. 查看目标实例的运行状态、网络信息等，其具体的含义，请参见下表所示。

☐	实例 ID / 名称	状态 ▾	网络	节点数量	规格与配置	连接地址	操作
☐	vdt- vdt-	运行中		1节点	计算型 2 核 4 GB	内网 http:// 外网 开启	登录 管理 更多 ▾
☐	vdb-eon1uhwm vdb-eon1uhwm	运行中		1节点	基础型 1 核 1 GB	内网 http:// 外网 开启	登录 管理 更多 ▾

参数	参数含义
实例 ID / 名称	● 每一个实例系统会随机分配一个 ID 编号。 ● 单击蓝色字体的实例 ID，可进入实例详情页面。 ● 实例名称为新建实例时设置的实例名称。
状态	显示当前数据库实例的运行状态。若为 运行中 ，说明实例运行正常。
网络	实例的私有网络与子网。私有网络与其子网之间使用-连接，例如：vectortest-sub1。单击 网络名称 ，进入私有网络控制台页面，可查看私有网络的详细信息。
节点数量	显示数据库集群当前的节点数量。

规格与配置	显示数据库实例的规格，具体规格所对应支持的向量维度与向量数量，请参见 产品规格 。
连接地址	显示给实例所分配的内网 IP 地址与端口。
操作	管理数据库实例。 ● 单击管理，进入 实例详情 页面。 ● 单击更多 > 退货/退费，可退还实例资源。具体操作，请参见 自助退还实例。 ● 单击更多 > 编辑标签，可在编辑标签窗口，设置标签键与标签值。具体操作，请参见 标签管理。 ● 单击更多 > 安全组，跳转至安全组管理页面。具体操作，请参见 管理安全组。

查看实例详情

在实例列表中，单击实例 ID，或在其操作列，单击管理，进入实例详情页面。

实例详情实例监控密钥管理安全组AI套件Embedding可视化管理

基本信息

实例 ID

vdb-

实例名称

vdb 

状态

运行中

地域

华南地区（广州）

可用区

两可用区

配置信息

计费模式

按量计费

创建时间

2025-03-08 13:03:49

到期时间

--

标签

规格信息

实例类型

高可用版

数据库版本

2.3

节点类型

计算型

节点规格

1 核 2 GB 

节点数量

2

磁盘容量

单节点磁盘容量20GB，总磁盘容量40GB

磁盘已使用/总容量

--/40GB

网络信息

私有网络



内网 IP

10.

端口

80 

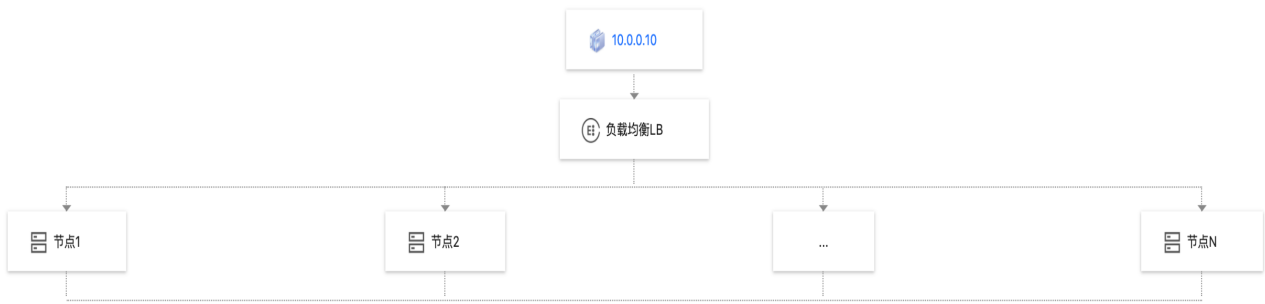
访问地址

http:// [连接并写入数据库](#) 

外网地址

[开启](#)

架构图



页面区域	参数名称	参数含义
基本信息	实例 ID	● 系统为实例随机分配一个 ID 编号。

		单击, 可直接复制实例 ID。
	实例名称	- 创建实例时, 为实例所配置的名称。 - 若未配置, 此处显示为空。
	状态	实例当前运行状态。 若为 运行中 , 说明该实例当前运行正常。
	地域	实例所属地域。
	可用区	所属可用区。
配置信息	计费方式	显示数据库实例的计费方式。
	创建时间	显示数据库实例的创建时间。
	到期时间	显示数据库公测结束时间。
	标签	显示数据库的标签信息。
规格信息	实例类型	显示实例当前的类型。 - 高可用版: 一主多从的分布式架构, 适合需要保证系统高可用性和容错性的大型企业或关键业务场景。 - 单机版: 适用于对高可用性和容错性需求不高的场景, 推荐个人、小型企业或测试/开发环境使用。
	实例类型	支持 单机版 与 高可用版 两类实例类型, 适用于不同的高可用性要求场景。具体信息, 请参见 选择实例类型与副本数 。
	数据库版本	数据库版本迭代信息。具体信息, 请参见 产品动态 。
	节点类型	显示当期数据库集群中的节点类型。产品实例类型信息, 请参见 产品规格 。
	节点规格	显示当期数据库集群中的节点规格。产品所支持的规格, 请参见 产品规格 。
	节点数量	显示当前数据库集群中所有的节点数量。
	磁盘容量	显示单节点的磁盘容量, 以及整个集群的磁盘容量。
	磁盘已使用/总容量	当前集群中所有存储节点已经使用的磁盘容量, 以及整个集群的磁盘总容量。
网络信息	私有网络	为实例配置的私有网络名称。
	内网 IP	为实例分配的腾讯云内网 IP 地址。

	端口	网络端口。
	访问地址	- 由内网 IP 与端口拼接成的 URL 地址，便于在连接实例时，可以快速复制访问地址。 - 单击连接并写入数据，可参考官网文档指引，连接实例，创建数据库，写入数据，检索数据。
	外网地址	开启外网地址，则显示外网连接数据库的地址。具体操作，请参见 开启外网访问 。
架构图	数据库集群的逻辑拓扑图。	

自助退还实例

最近更新时间：2024-12-26 15:15:02

操作场景

- 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**。
- 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。

❗ 说明：

自助退还实例至回收站之后，实例的状态将变为**已隔离**或**待删除**时，不再产生与该实例相关的费用。

退费说明

包年包月实例销毁后，**退款处理原则**如下：

- 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。
- 自助退还计算规则及退费示例，更多信息，请参见 [退费说明](#)。

操作步骤

- 使用腾讯云账号登录 [向量数据库控制台](#)。
- 在页面上方，**向量数据库 - 实例列表**右边的下拉框选择地域。
- 在实例列表上方右上角的搜索框中，输入实例名称模糊搜索目标实例，找到需退还的实例，在其**操作列**选择**更多 > 退货/退费**。
- 在**退货退费**的小窗口，请阅读退还实例注意事项，并勾选**已阅读并同意**，单击**立即销毁**。

退货退费

×

已选择销毁1个实例

实例 ID	实例名称
vdb-	折扣

i

实例退货并彻底销毁后数据将无法找回, 请提前备份数据
 实例彻底销毁后IP资源同时释放
 实例彻底销毁后,退款处理如下:

- 5天无理由自助退款的金额将退还至原支付账户
- 普通自助退款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号。详细规则请查看XXX

☒ 已阅读并同意

立即销毁

取消

- 在退订清单中，确认退订信息与退费金额，单击**确认退订**，在弹出的小窗口，再次单击**确认退订**，完成退费操作。
- 实例当前状态变更为**隔离中**，等待任务执行完成，实例将被移除实例列表，已完成退还。在左侧导航栏选择**回收站**，可看到实例已经被隔离在回收站中。

后续操作

退还实例在回收站中已隔离，在保留时长内，还可恢复。具体操作，请参见 [恢复已隔离实例](#)。

恢复已隔离实例

最近更新时间：2025-06-11 15:26:32

操作场景

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。具体回收机制，如下所示。

包年包月实例进入回收站

- **保留时长：**回收站内实例保留15个自然日。
- **过期处理：**15个自然日后未进行续费，系统将释放资源，所有数据将被清除且不可恢复。

❗ 说明：

云服务资源到期前七天，系统会开始给用户发送续费提醒通知，到期后第八天开始，此云数据库不可再使用，将被回收至回收站中。

按量计费实例进入回收站

- **保留时长：**在未欠费的情况下，用户主动销毁的实例在回收站内保留15天。
- **过期处理：**没有按时续费的实例，超过保留时长后，系统将释放实例资源，不可恢复。

⚠ 注意：

- 账户余额为0后，24小时后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。
- 已进入回收站的按量计费实例，账户欠费时不能进行恢复操作，请您先进行续费操作。
- 由于按量计费实例最长在回收站保存15天，请您留意释放时间，请及时续费和恢复。

前提条件

- 实例已被退还隔离在回收站中。
- 腾讯云账户余额充足。

恢复单个实例

1. 使用腾讯云账号登录 [向量数据库控制台](#)。
2. 在左侧导航栏，选择[向量数据库](#) > [回收站](#)，在回收站的实例列表中，可看到已经被隔离的所有实例。
3. 在列表右上方的输入框，输入实例名称或 ID 搜索目标实例，找到需恢复的实例。
4. 在目标实例的操作列，根据实例的不同计费方式，选择不同操作，恢复实例。
 - **包年包月实例**：单击**续费**，在续费小窗口，执行如下操作，支付续费金额，即可恢复实例。
 - 单击**查看详情**，确认实例信息。
 - 在**续费时长**，选择实例需续费的时长，按整月或整年选择。
 - 在**配置费用**后面，确认实例续费的总金额。



续费窗口截图，显示续费时长选择为1个月，配置费用为0.00元。窗口底部有“确定”和“取消”按钮。

续费

确定对所选实例进行续费?已选择4个实例 [查看详情](#) ▼

续费时长 ☒ 1个月 ☐ 2个月 ☐ 3个月 ☐ 6个月 ☐ 1年 ☐ 2年 ☐ 3年 ☐ 其他时长

配置费用

☒ 我已阅读并同意《[云数据库服务条款](#)》、《[服务等级协议](#)》和《[退费说明](#)》

- **按量计费实例**：单击**开机**，在**实例恢复**小窗口，确认实例信息，单击**确定**，即可恢复实例。



实例恢复窗口截图，显示已选择1个实例，实例ID为vdb-，实例名称为。窗口底部有“确定”和“取消”按钮。

实例恢复

已选择1个实例

实例 ID	实例名称
vdb-	

批量恢复包年包月实例

说明：

回收站已隔离实例，计费模式为按量计费，不支持批量恢复操作，仅包年包月支持。

1. 使用腾讯云账号登录 [向量数据库控制台](#)。
2. 在左侧导航栏，选择[向量数据库](#) > [回收站](#)，在回收站的实例列表中，可看到已经被隔离的所有实例。
3. 在列表中，勾选需恢复的多个包年包月的实例。

恢复							
可通过实例 ID/实例名称搜索							
实例 ID / 名称	状态	网络	节点数量	规格与配置	连接地址	计费模式	操作
☒ vdb- vdt-	已隔离		1节点	存储型 1 核 8 GB	内网 http:// 外网 开启	包年包月 到期时间: 已到期2天	续费 立即下线
☒ vdb- vdt-	已隔离		1节点	存储型 1 核 8 GB	内网 http:// 外网 开启	包年包月 到期时间: 已到期2天	续费 立即下线
☒ vdb- vdt-	已隔离		1节点	存储型 1 核 8 GB	内网 http:// 外网 开启	包年包月 到期时间: 已到期6天	续费 立即下线
☒ vdb- vdt-	已隔离		1节点	计算型 2 核 4 GB	内网 http:// 外网 开启	包年包月 到期时间: 已到期3天	续费 立即下线

4. 单击列表上方的恢复，在续费小窗口，在续费时长后面选择包年包月的时长，在配置费用后面确认付费金额，勾选我已阅读并同意《云数据库服务条款》、《服务等级协议》和《退费说明》，单击确定，支付费用，完成续费，便可以直接恢复实例。

续费

确定对所选实例进行续费? 已选择4个实例 [查看详情](#)

续费时长

1个月

2个月3个月6个月1年2年3年其他时长

配置费用

☒ 我已阅读并同意《[云数据库服务条款](#)》、《[服务等级协议](#)》和《[退费说明](#)》

确定取消

5. 在实例列表中，确认实例已从回收站中恢复至实例列表。

立即下线

最近更新时间：2024-05-13 15:56:42

操作场景

当一个项目完成或者不再需要实例资源时，或者有些资源因成本控制等因素需要进行及时清理时，可以立即删除相关的实例，释放资源，避免闲置已隔离实例占用不必要的资源。

⚠ 注意：

实例退货并彻底销毁后数据将无法找回，IP 资源同时释放，请提前备份实例数据。

操作步骤

1. 使用腾讯云账号登录 [向量数据库控制台](#)。
2. 在页面上方，**向量数据库 - 实例列表**右边的下拉框选择地域。
3. 在**实例列表**上方右上角的搜索框中，输入实例名称模糊搜索目标实例，找到需删除的实例。
4. 在其**操作列**选择**更多 > 退货/退费**，退还实例。具体操作，请参见 [自助退还实例](#)。
5. 在左侧导航栏，选择**向量数据库 > 回收站**，在回收站的实例列表中，可看到已经被隔离的所有实例。
6. 在列表右上方的输入框，输入实例名称或 ID 搜索目标实例，找到需删除的实例。
7. 在其**操作列**，单击**立即下线**。

☐ 实例 ID / 名称	状态 ▾	网络	节点数量	规格与配置	连接地址	计费模式	操作
☐ [实例名称]	已隔离	vcl [网络]	1节点	存储型 1 核 8 GB	内网 http:// [地址] 外网 开启 [地址]	包年包月 到期时间：已到期2天	续费 立即下线
☐ [实例名称]	已隔离	vcl [网络]	1节点	存储型 1 核 8 GB	内网 http:// [地址] 外网 开启 [地址]	包年包月 到期时间：已到期3天	续费 立即下线

8. 在实例下线小窗口，确认下线实例的 ID 与名称，单击**确定**，完成删除操作。

标签管理

最近更新时间：2024-05-13 15:40:21

腾讯云向量数据库（Tencent Cloud VectorDB）支持在控制台编辑实例标签，便于您通过标签管理实例。

背景信息

标签由标签键和标签值组成，可以标记腾讯云向量数据库实例。如果您的腾讯云账号下有多种云资源，不同的资源类型之间有多种关联，且云上资源日益增加，管理难度也随之变化。您可以通过标签将作用相同或者相关联的资源进行分组归类。日常运维或者定位问题时，您便可以根据标签快速检索资源，进行批量操作，高效运维。

计费说明

标签管理是腾讯云为您的腾讯云账户提供的免费服务之一，不另行收费。可直接进入 [控制台](#) 使用产品。

使用须知

- 1个标签包含1个标签键和1个标签值（tagKey:tagValue）。
- 1个实例最多可以绑定50个标签。
- 1个实例上的同一个标签键只能对应1个标签值。

前提条件

已创建实例。

操作步骤

1. 使用腾讯云账号登录 [向量数据库控制台](#)。
2. 在实例列表中，找到需编辑标签的实例。
3. 在目标实例操作列，选择**更多** > **编辑标签**，进入编辑标签窗口。
4. 在编辑标签窗口，在**标签键**下拉列表中重新选择合适的标签键，并在**标签值**输入框选择对应的标签值。单击**添加**，可添加多条标签。

编辑标签

×

① 编辑须知

- 标签用于从不同维度对资源分类管理。如现有标签不符合您的要求，请前往 [标签管理](#)。

已选择 1 个资源

标签键

▼

标签值

▼

×

+ 添加

↻ 键值粘贴板

确定

取消

5.（可选）如果现有标签不符合您的业务要求，请执行以下操作：

- 在当前页面右上角，单击**标签管理**。
- 在**标签管理**页面，单击**新建标签**。
- 在**新建标签**页面，认真了解设置标签的注意信息。
- 在**标签键**输入框，设置新的标签值，在**标签值**输入框，输入对应的标签值。标签键设置要求如下：
 - 字符长度要求[1,63]。
 - 可输入英文字母、数字以及中文字符。
 - 可输入特殊符号：加号“+”、等号“=”、下划线“_”、短划线“-”、英文圆点“.”、英文冒号“:”、斜线“/”、at“@”、英文括号“()”、中文括号“（）”、方括号“[]”、中括号“【】”。
- 单击**确定**，完成创建。
- 再返回数据库实例的**编辑标签**页面，在标签键的下拉列表中，单击**重新加载**，可以选择新建的标签键，再选择对应的标签值。

6. 单击**确定**，完成设置。

更多参考

关于标签管理的更多信息，请参见 [标签管理](#)。

配置变更

升配节点规格

最近更新时间：2025-04-23 16:55:01

节点规格升级是指调整节点的 CPU 或内存到更高的配置。如果您在实例监控页面发现节点的 CPU 或内存使用率较高（例如，CPU 或内存使用率持续接近80%），且有更高性能或更大数据量存储需求，建议升级节点的 CPU 或内存配置以更好支持业务发展，避免因 CPU 资源不足而不能支持更高性能，或内存写满后不能继续写入给业务造成影响。

前提条件

- 确保实例当前运行稳定，状态为**运行中**，没有正在进行的重大操作。
- 检查实例当前资源使用情况（CPU、内存、磁盘容量），评估目标规格满足业务需求。

❗ 说明：

升配节点规格需要满足如下要求。若不满足，请 [提交工单](#) 申请。

- 实例类型要求为**高可用版三可用区**，数据库版本**大于等于2.4**。请在控制台**实例详情**页面的**规格信息**区域确认实例类型与数据库版本。
- 实例**副本数均大于等于2**。请通过 </collection/describe> 查看 collection 信息的副本数量。

注意事项

❗ 说明：

- 为确保系统升级规格期间的数据完整性与服务稳定性，DDL 操作、索引操作及 AI 套件相关操作将被临时限制，除此之外读写可正常提供服务，请在操作完成后再执行相关变更。
- 在节点规格扩容过程中，可能会遇到以下情况：客户端连接闪断，强一致性读操作间歇性报错，建议您提前评估业务容忍度，并配置完善的重试机制，尽可能选择在业务低峰期进行扩容操作，以确保业务的连续性和稳定性。

操作步骤

1. 使用腾讯云账号登录 [向量数据库控制台](#)。
2. 在页面上方，**向量数据库 - 实例列表**右边的下拉框选择地域。
3. 在实例列表右上角的搜索框中，输入实例名称模糊搜索目标实例，找到需变更配置的实例，在其**操作列**选择**配置变更 > 升配节点规格**，进入如下页面。

升配节点规格

×

❶ 1. 配置变更后，实例将按照新的规格计费

2. 在单次配置变更中，CPU、内存只能同时升配

实例ID/名称 vdb-

地域 ap-chengdu

实例类型 高可用版

可用区 两可用区

磁盘已使用/总容量 3.42GB/180GB

内存已使用/总容量 3.42GB/180GB

节点类型 计算型 ▾

节点规格 1核2G ▾

节点硬盘容量

10 GB

1000 GB

–

60

+

GB

 (步长为: 10GB)

对比	节点类型	节点数	单节规格	总内存容量(GB)	总磁盘容量(GB)
当前配置	计算型	3	1 核 2 GB	6	180
新配置	计算型	3	1 核 2 GB	6	180

配置费用 元 /小时

操作时间

维护时间内

 立即调整

维护时间00:00-03:00(实例详情页可修改维护时间)

确定

取消

4. 确认升配实例的信息，根据下表配置相关参数。

界面参数	解释说明
节点类型	默认展示当前节点类型，若需变更，可选择变更后的节点类型，支持选项：计算型、标准型、存储型。更多节点类型的信息，请参见 选择实例类型与副本数 。
节点规格	● 默认展示当前规格的下一个层级的规格，例如：当前8核16G，则展示为16核32G。 ● 如果变更节点类型，比如从计算型调整为存储型，节点规格可选项（CPU & 内存）必须 >= 当前配置。 ❗ 说明： 为了确保业务的稳定运行，建议定期评估业务数据的增长趋势，选择一个与业务需求相匹配的节点规格。如果对选择最佳规格不太确定，可以逐级提升

	规格，直至达到理想的性能和成本平衡点。
节点硬盘容量	默认展示当前磁盘容量。滑动选择器最小值为当前磁盘容量值，最大值为1000GB，步长10GB。
对比	使用表格形式展示新规格与旧规格关键数据（节点类型、节点数、节点规格、总内存、总磁盘容量）的对比差异，便于直观地确认变更数据。
配置费用	配置变更后的费用。 ● 按量计费：实例调整配置后每小时的计费单价，在每个整点（北京时间）进行一次费用结算。 ● 包年包月：实例调整配置后，在到期时间内，还需支付的费用。
操作时间	● 维护时间：在下一个维护时间段内执行切换实例规格任务。关于维护时间的更多信息，请参 维护时间窗。 ● 立即调整：立即执行调整实例规格任务，切换新规格。调整实例内存与磁盘可能涉及节点迁移或者主从切换，主从切换时间点将不可控，可能导致断连或重启，请确保此时没有重大业务操作。

5. 单击**确定**，升级实例配置，等待任务执行完成，在**实例详情**页面的**规格信息**区域，可看到升级后的节点规格。

增加节点数量

最近更新时间：2025-04-18 17:58:11

操作场景

单节点规格已达上限（例如，节点为计算型，32核64GB），无法再扩容节点规格至更高的配置时，如期望进一步提升实例的整体性能，可以通过增加节点数量来实现横向扩展。

前提条件

- 确保实例当前运行稳定，状态为**运行中**，没有正在进行的重大操作。
- 检查实例当前资源使用情况（CPU、内存、磁盘容量），评估目标规格满足业务需求。

操作步骤

注意：

增加节点数量（水平扩容）过程中，由于节点间数据搬迁可能导致请求处理耗时暂时升高，建议评估业务影响后再进行操作。

1. 使用腾讯云账号登录 [向量数据库控制台](#)。
2. 在页面上方，**向量数据库 - 实例列表**右边的下拉框选择地域。
3. 在实例列表右上角的搜索框中，输入实例名称模糊搜索目标实例，找到需变更配置的实例，在其**操作列**选择**配置变更 > 增加节点数**，进入如下页面。

增加节点数

1. 配置变更后，实例将按照新的规格计费

2. 在单次配置变更中，CPU、内存只能同时升配

实例ID/名称 vdb-

地域 ap-chengdu

实例类型 高可用版

可用区 三可用区

磁盘已使用/总容量 --/40GB

内存已使用/总容量 --/40GB

节点数

-

2

+

支持选择 - 30节点

对比	节点类型	节点数	单节规格	总内存容量(GB)	总磁盘容量(GB)
当前配置	计算型	2	1核 2 GB	4	40
新配置	计算型	2	1核 2 GB	4	40

配置费用

元

操作时间

维护时间内

立即调整

维护时间00:00-03:00(实例详情页可修改维护时间)

确定

取消

4. 确认升配实例的信息，根据下表配置相关节点参数。

界面参数	解释说明
节点数	默认展示当前实例的节点数，最小可选为当前节点数，最大为30。
对比	使用表格形式展示新规格与旧规格关键数据（节点类型、节点数、节点规格、总内存、总磁盘容量）的对比差异，便于直观地确认变更数据。
配置费用	配置变更后的费用。 按量计费：实例调整配置后每小时的计费单价，在每个整点（北京时间）进行一次费用结算。 包年包月：实例调整配置后，在到期时间内，还需支付的费用。

操作时间	● 维护时间内：指在下一个维护时间段内执行增加节点数的任务。实例列表中“状态”列将显示“待执行配置变更”，等到维护时间窗时，启动扩容任务。维护时间的更多信息，请参 维护时间窗。 ● 立即调整：立即执行增加节点数量的任务。请确保此时没有重大业务操作。
------	---

5. 单击**确定**。

维护时间窗

最近更新时间：2025-04-18 17:58:11

操作场景

为保证向量数据库实例的稳定性，后台系统会不定期在维护时间内对实例进行维护操作。建议您对业务实例设置自己可接受的维护时间，一般设置在业务低峰期，将对业务的影响降到最低。通常，以扩容节点规格为例，当选择操作时间为维护时间窗内时，后台会在下一个维护时间自动进行扩容操作，降低其他时间执行时对业务的影响。

设置维护时间窗

1. 使用腾讯云账号登录 [向量数据库控制台](#)。
2. 在页面上方，**向量数据库 - 实例列表**右边的下拉框选择地域。
3. 在实例列表右上角的搜索框中，输入实例名称模糊搜索目标实例，找到目标实例。
4. 单击**实例 ID**，进入**实例详情**页面。
5. 在**配置信息**区域，单击**维护时间窗**后面的.

配置信息

计费模式	按量计费
创建时间	2025-03-07 09:03:56
到期时间	--
维护时间窗	00:00-03:00 
标签	 

6. 在**修改维护时间窗**的小窗口，设置**维护开始时间**，并选择**维护窗口时长**，确认**维护窗口范围**，单击**确定**。

修改维护时间窗 ×

维护开始时间 *

00:00 

维护窗口时长 *

3 小时 

维护窗口范围

00:00 - 03:00

确定

取消

7. 单击**确定**，完成操作。

监控实例

查看监控指标

最近更新时间：2024-12-17 16:26:02

腾讯云向量数据库（Tencent Cloud VectorDB）支持通过腾讯云可观测平台（Tencent Cloud Observability Platform，TCOP）实时监控实例资源的监控指标数据，辅助运维人员快速识别并解决潜在的性能问题，保障业务稳定运行。

监控指标

监控系统从实例与节点两个层面实时跟踪数据库的关键性能指标，包含资源使用、流量、请求响应等。

- 实例维度监控：监控覆盖整个数据库集群，包括集群的 CPU、内存和磁盘的最大与平均使用率，以及数据库出入流量和请求响应情况，以便全面评估整个实例的性能状态。
- 节点维度监控：监控聚焦于每一个节点的 CPU、内存和磁盘的使用率，以及数据流量与请求响应情况，以便快速识别出资源处理瓶颈，预防内存溢出或性能下降。

监控对象	分组	指标名称	指标英文名称	单位	含义
实例监控	CPU 监控	节点 CPU 最大使用率	cluster_nodes_max_cpu_util	%	集群中单个节点的最高 CPU 使用率。
		节点 CPU 平均使用率	cluster_nodes_avg_cpu_util	%	集群中所有节点的平均 CPU 使用率。
	内存 监控	节点内存最大使用率	cluster_nodes_max_mem_util	%	集群中单个节点的最高内存使用率。
		节点内存平均使用率	cluster_nodes_avg_mem_util	%	集群中所有节点的平均内存使用率。
	磁盘 监控	节点磁盘最大使用率	cluster_nodes_max_disk_usage_util	%	集群中单个节点的最高磁盘使用率。
		节点磁盘平均使用率	cluster_nodes_avg_disk_usage_util	%	集群中所有节点的平均磁盘使用率。

	网络 监控	集群网络 入流量	cluster_nodes _total_in_flow	MBy tes/ s	每秒流入数据库的流量。
		集群网络 出流量	cluster_nodes _total_out_flo w	MBy tes/ s	每秒流出数据库的流量。
	请求 监控	集群插入 请求平均 耗时	cluster_nodes _avg_upsert_t ime	ms	集群处理插入请求所需的平均时 间。
		集群请求 异常率	cluster_nodes _commands_ except_util	%	集群处理请求异常数与总请求数之 比。
		集群请求 成功率	cluster_nodes _commands_ success_util	%	集群处理请求成功数与总请求数之 比。
		集群请求 超时率	cluster_nodes _commands_t imeout_util	%	集群处理请求超时的数量与总请求 数之比。
		集群请求 量QPS	cluster_nodes _total_comma nds	Cou nt/s	集群每秒总请求数。
		集群请求 异常数	cluster_nodes _total_excepti on_command s_cnt	Cou nt/s	集群每秒请求异常数。
		集群搜索 请求量 QPS	cluster_nodes _total_search _cnt	Cou nt/s	集群每秒搜索请求。
		集群请求 超时数	cluster_nodes _total_timeout _commands_ cnt	Cou nt	集群每秒请求超时数。
		集群插入 请求量 QPS	cluster_nodes _total_upsert _cnt	Cou nt/s	集群每秒插入请求数。

节点监控	响应监控	集群请求平均耗时	cluster_nodes_avg_commands_time	ms	集群处理请求的平均耗时。
		集群搜索请求平均耗时	cluster_nodes_avg_search_time	ms	集群处理搜索请求所需的平均时间。
		空结果率	cluster_nodes_avg_null_result_util	%	集群各节点空结果返回请求数与各节点总请求量之和的比例。
		集群搜索平均召回数	cluster_nodes_avg_recall_count	Count	集群处理搜索请求的平均召回数量，通过各节点平均召回数之和除以节点数计算平均召回数。
	CPU 监控	CPU 使用率	cpu_util	%	节点 CPU 使用率
	内存监控	内存使用率	mem_util	%	实际使用内存和申请总内存之比
	磁盘监控	硬盘空间使用率	disk_util	%	实际使用硬盘容量和节点总容量之比
		硬盘空间使用量	disk_used	MBytes	实际使用硬盘容量
	网络监控	网络入流量	in_flow	MBytes/s	内网每秒所流入的字节数
		网络出流量	out_flow	MBytes/s	内网每秒所流出的字节数
	请求监控	请求量 QPS	commands	Count/s	每秒请求总数量
		请求成功率	commands_success_util	%	请求成功数量与总请求数之比
		请求超时率	commands_timeout_util	%	超时请求数与总请求数之比
		请求异常率	commands_exp_util	%	请求异常数与总请求数之比

		失败请求 量QPS	commands_fa il_cnt	Cou nt/s	每秒无法完成的请求量
		请求平均 耗时	commands_ti me	ms	从发起请求到接收到请求所耗费的时 间的平均值
		搜索请求 数量 QPS	search_cnt	Cou nt/s	每秒搜索请求的总数量
	响应 监控	空结果率	search_null_r es_util	%	搜索返回结果为空的数量与搜索请 求量之比
		平均召回 数	search_recall _avg	Cou nt	搜索返回的文档数与搜索批量之比

前提条件

- 已 [创建数据库实例](#)。
- 实例状态为**运行中**。

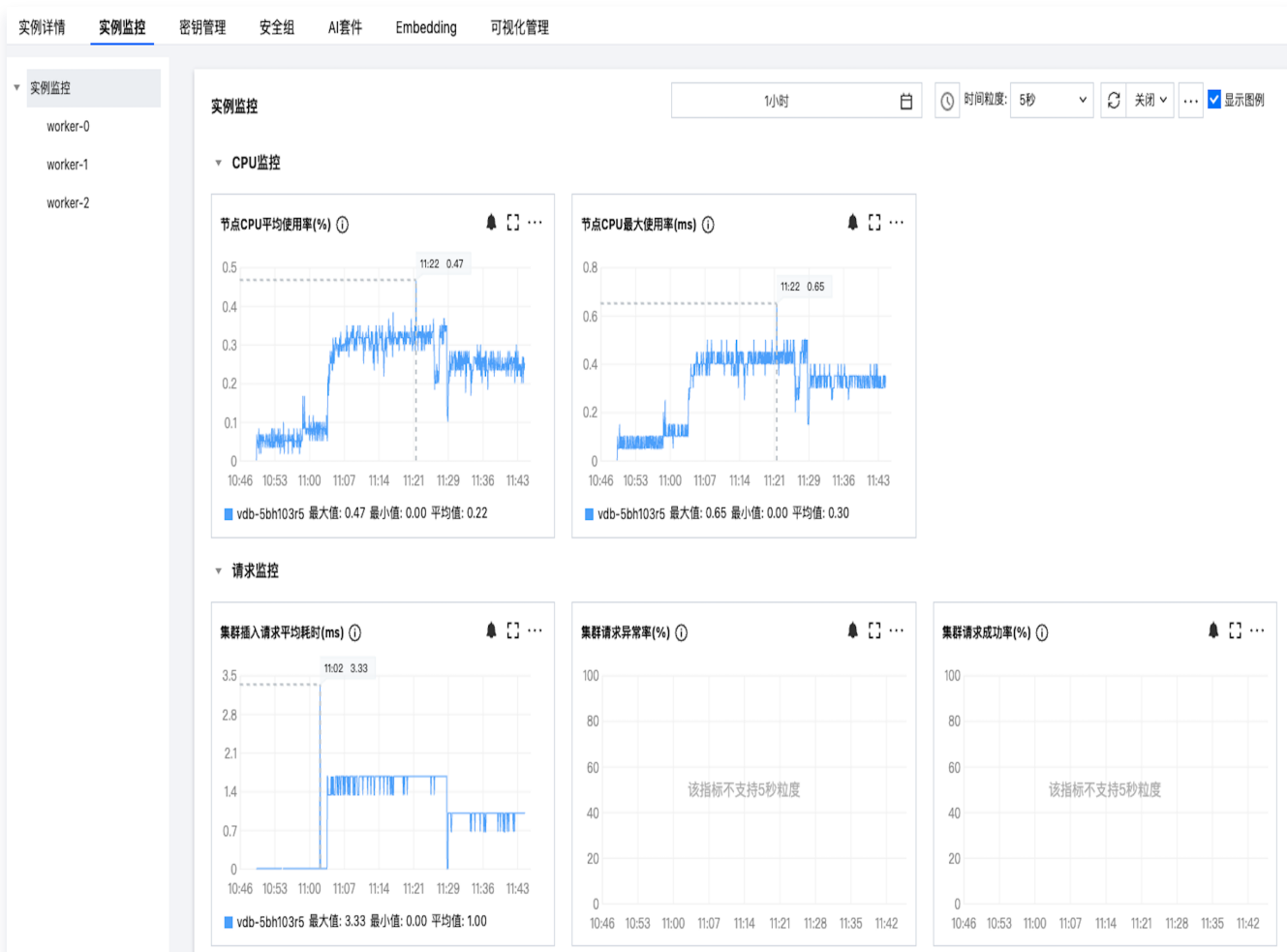
操作指南

1. 使用腾讯云账号登录 [向量数据库控制台](#)。
2. 在实例列表中，找到需查看监控数据的实例。
3. 单击目标**实例 ID**，或在其**操作列**，单击**管理**，进入**实例详情**页面。
4. 切换至**实例监控**页签，可直接查看实例维度的监控视图；选择**实例监控**下的节点：worker-0、worker-1 或 worker-2，可分别查看单个节点的监控视图。
 - 在页面右上角的时间框，选择所需监控数据的时间跨度。
 - 在页面右上角的**时间粒度**后面的下拉列表选择监控粒度，即监控数据的采集粒度。时间跨度的范围不同，所支持的时间粒度不同，如下表所示。

时间跨度	时间粒度	监控数据保留时长
5分钟	5秒、10秒、1分钟	时间粒度不同，监控数据存储保留时长不同。 ● 秒级：1天。 ● 1分钟：15天。 ● 5分钟：31天。 ● 1小时：93天。 ● 1天：186天。
30分钟	5秒、10秒、1分钟与5分钟	
1小时		
3小时	5秒、10秒、1分钟、5分钟、1小时	
12小时	1分钟、5分钟、1小时	

1天（24小时）	1分钟、5分钟、1小时、1天
2天	
7天	1小时、1天
30天	1小时、1天

- 单击 ，可刷新监控数据，并更新监控视图。
- 单击关闭，可关闭数据更新；在关闭的下拉菜单，可设置更新数据的周期，支持每隔30s、5min、30min、1h更新数据。



5. 如果需导出监控视图的监控数据与视图，请在每一个视图的右上角单击 ，在其下拉列表，可选择数据导出与图片导出。



6. 如果需全屏展示单个监控视图，单击 ；如果需给单个监控指标配置告警，单击 ，具体操作，请参见 [配置监控告警](#)。

配置监控告警

最近更新时间：2024-05-13 15:46:01

操作场景

为了防止某些监控指标达到一定值后，影响您系统的正常运行。您可以对这些监控指标设定告警规则促使告警系统自动检查监控数据，并在监控数据满足条件时，发送告警通知给管理员，帮助您第一时间了解业务异常，并迅速解决。

背景信息

腾讯云可观测平台（Tencent Cloud Observability Platform，TCOP）是一项可对云产品资源实时监控和告警的服务。腾讯云向量数据库（Tencent Cloud VectorDB）通过腾讯云可观测平台提供对监控指标的告警功能。告警在一定周期内监控某些特定指标，并根据给定的阈值，每隔若干个时间段通过多种方式（微信、短信等）发送告警通知。更多信息，请参见 [腾讯云可观测平台告警管理](#)。

计费说明

- 腾讯云可观测平台配置告警策略，以监控实例各项关键指标，可免费使用。
- 目前只针对告警短信、电话告警收费，具体信息，请参见 [腾讯云可观测平台计费概述](#)。

前提条件

- 开通 [腾讯云可观测平台](#) 服务。
- 数据库实例状态为运行中。
- 已收集告警通知对象的信息，包括：邮件、短信、电话等。

操作步骤

1. 使用腾讯云账号登录 [向量数据库控制台](#)。
2. 在实例列表中，找到需配置告警规则的实例。
3. 单击目标实例 ID，或在其操作列，单击管理，进入实例详情页面。
4. 在实例详情页面，选择实例监控，可查看到实例的监控视图。
5. 在每一个监控视图的右上角，单击 ，进入新建告警策略页面。
6. 在新建告警策略页面，请参见下表，配置告警策略。告警策略的基本概念，请参见 [新建告警策略](#)。

← 新建告警策略

1 配置告警

>

2 配置告警通知

基本信息

策略名称

最多60个字符

备注

最多100个字符

配置告警规则

监控类型

云产品监控

应用性能监控HOT

前端性能监控HOT

云拨测HOT

策略类型

云数据库 / 向量数据库 / 节点

已有 0 条，还可以创建 300 条静态阈值策略；当前账户有0条动态阈值策略，还可创建20条。

所属标签

标签键

标签值

×

+ 添加

键值粘贴板

告警对象

实例ID

1个vdl

触发条件

☐ 选择模板

☒ 手动配置

指标告警

满足以下

所有

指标判断条件时，触发告警

☐ 启用告警分级功能

阈值类型

☒ 静态

☐ 动态

if

CPU使用率

统计粒度1分钟

>

0

%

持续 1 个数据点

🗑

阈值类型

☒ 静态

☐ 动态

if

CPU使用量

统计粒度1分钟

>

0

%

持续 1 个数据点

🗑

添加指标

then

每2小时告警一次

📘

上一步

下一步：配置告警通知

界面区域	参数名称	参数解释
基本信息	策略名称	给告警策略自定义名称，便于识别即可。
	备注	简要描述告警策略，便于识别。

	监控类型	请选择云产品监控。
	策略类型	可设置的策略类型：云数据库 / 向量数据库 / 节点。
	所属标签	请分别配置告警策略的标签，便于通过标签管理告警策略。
配置告警规则	告警对象	● 选择实例 ID：则该告警策略绑定指定的数据库实例。 ● 选择实例分组：则该告警策略绑定指定的数据库实例组。如何创建实例组，请参见 实例分组。 ● 选择全部对象：则该告警策略绑定当前账号拥有权限的全部实例。
	触发条件	● 选择模板：在下拉列表选择模板文件，将根据模板文件预置的触发条件上报告警。具体配置，请参阅 配置触发条件模板。 ● 手动配置：需在下方指标告警区域，逐一配置每一条指标触发告警的阈值条件。指标告警区域的阈值类型，可选择静态和动态。更多信息，请参见 新建告警策略。
配置告警通知	通知模板	支持选择系统预设通知模板和用户自定义通知模板，每个告警策略最多只能绑定三个通知模板。具体信息，请参见 通知模板。 ● 选择模板：如果已预设通知模板，单击选择模板，在选择通知模板窗口，选择已经创建的告警通知模板，单击确定。 ● 新建模板：单击新建模板，在新建通知模板窗口，设置通知模板名称，通知周期、接收对象等信息，单击确定。

7. 单击**完成**，等待任务执行完成，在**告警管理**页面的**策略管理**页签，可查看到已经创建的告警策略。

管理密钥

最近更新时间：2024-06-26 12:36:11

API 密钥是一种用于身份验证的机制，用于标识和验证 API 请求的发送者。在发送 API 请求时，您需要在请求头中携带密钥 ID 和签名信息，以便腾讯云可以验证请求的合法性。API 密钥是保护您的账户和数据安全的重要机制。

前提条件

- 已 [创建数据库实例](#)。
- 实例状态为 **运行中**。

操作指南

1. 使用腾讯云账号登录 [向量数据库控制台](#)。
2. 在实例列表中，找到目标实例。
3. 单击目标实例 ID，或在其操作列，单击**管理**，进入**实例详情**页面。
4. 在**实例详情**页面，选择**密钥管理**。
5. 在**密钥管理**页面，可查看到访问数据库的账户及其对应的 **API 密钥**。

实例详情 实例监控 <u>密钥管理</u> 安全组 AI套件 Embedding 可视化管理		
名称	API 密钥	操作
root	***** 显示 复制	重置密钥

6. 查看并复制 API 密钥信息以及更新密钥。
 - 查看API 密钥：单击 **API 密钥** 列的显示，可查看 API 密钥信息，并进行复制。
 - 复制密钥：单击**API 密钥**列的复制，复制密钥信息。
 - 重置密钥：单击**操作**列的**重置密钥**，对密钥进行更新。

管理安全组

最近更新时间：2024-06-25 09:55:41

背景信息

安全组 是一种有状态的包含过滤功能的虚拟防火墙，用于设置单台或多台云数据库的网络访问控制，是腾讯云提供的重要的网络安全隔离手段。安全组是一个逻辑上的分组，您可以将同一地域内具有相同网络安全隔离需求的云数据库实例加到同一个安全组内。云数据库与云服务器等共享安全组列表，安全组内基于规则匹配，具体规则与限制请参见 [安全组详细说明](#)。

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。

⚠ 注意：

- 云数据库安全组目前仅支持私有网络 VPC 内网访问的网络控制，暂不支持对基础网络的网络控制。
- 由于云数据库没有主动出站流量，因此出站规则对云数据库不生效。

操作步骤

步骤一：创建安全组

- 登录 [私有网络控制台](#)。
- 在左侧导航选择**安全组**页，在右侧页面上方选择地域，单击**新建**。
- 在弹出来的窗口中，完成如下配置，确认后单击**确定**。

新建安全组

×

模板

放通22, 80, 443, 3389端口和ICMP协议

名称

放通22, 80, 443, 3389端口和ICMP协议-20230331114

所属项目

默认项目

备注

公网放通云主机常用登录及web服务端口，内网全放通。

高级选项

显示模板规则

确定

取消

○ **模板**：在下拉列表选择安全组模板。

○ **放通全部端口**：默认放通全部端口到公网和内网，具有一定安全风险。安全组规则已默认添加。单击下方的**显示模板规则**，可查看该安全组模板的**出站规则**与**入站规则**。

○ **放通22, 80, 443, 3389端口和 ICMP 协议**：默认放通22, 80, 443, 3389端口和 ICMP 协议，内网全放通。安全组规则已默认添加。

○ **自定义**：安全组创建成功后，按需添加安全组规则。

○ **名称**：自定义设置安全组名称。

○ **所属项目**：默认选择“默认项目”，可指定为其他项目，便于后期管理。

○ **备注**：自定义，简短地描述安全组，便于后期管理。

○ **高级配置**：给安全组添加标签。

4. 如果模板为自定义，在**提醒**对话框，单击**立即设置规则**，执行以下步骤。

步骤二：设置安全组入站规则

1. 在**安全组规则**页面，选择**入站规则**页签，单击**添加规则**。

2. 在弹出**添加入站规则**的窗口中，设置规则。

添加入站规则

类型

来源 ①

协议端口 ①

策略

备注

自定义

如10.0.0.1或10.0.0.0/16

如UDP:53,TCP:80,443或TCF

允许

+新增一行

完成

取消

- **类型**：请选择默认类型自定义。
- **来源**：设置访问数据库的源，即入站来源。支持以下格式定义来源。

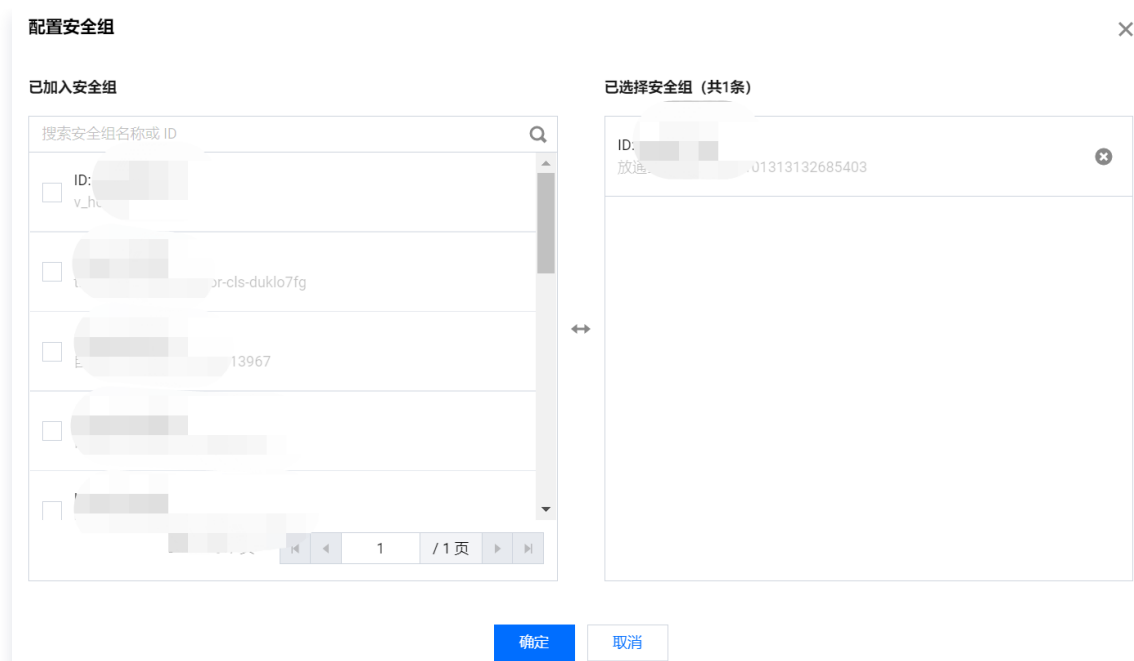
来源格式	格式说明
CIDR 表示	单个 IPv4 地址或 IPv4 地址范围用 CIDR 表示法（如203.0.113.0、203.0.113.0/24或者0.0.0.0/0，其中0.0.0.0/0代表匹配所有 IPv4 地址）。 单个 IPv6 地址或 IPv6 地址范围用 CIDR 表示法（如FF05::B5、FF05:B5::/60、::/0或者0::0/0，其中::/0或者0::0/0代表匹配所有 IPv6 地址）。
安全组 ID	引用安全组 ID 来匹配安全组关联的服务器的 IP 地址。
参数模板	引用 参数模板 中的 IP 地址对象或 IP 地址组对象。

- **协议端口**：填写客户端访问腾讯云向量数据库的协议类型和端口。您可在 实例列表 的内网地址列查看端口信息，默认为80。
- **策略**：默认选择“允许”。
 - 允许：放行该端口相应的访问请求。
 - 拒绝：直接丢弃数据包，不返回任何回应信息。
- **备注**：自定义，简短地描述规则，便于后期管理。

3. 单击**完成**，完成安全组入站规则的添加。

步骤三：给实例绑定安全组

- 使用腾讯云账号登录 向量数据库控制台。
- 在实例列表中，找到需绑定安全组的实例。
- 单击目标实例 ID，或在其操作列，单击**管理**，进入实例详情页面。
- 选择**安全组**页签，单击**配置安全组**。



5. 在配置安全组对话框，选择需要绑定的安全组，单击确定。

更多操作

调整已绑定安全组的优先级

1. 使用腾讯云账号登录 [向量数据库控制台](#)。
2. 在实例列表中，找到需绑定安全组的实例。
3. 单击目标实例 ID，或在其操作列，单击管理，进入实例详情页面。
4. 选择安全组页签，可查看到实例当前所有的安全组。
5. 单击编辑，您可以在操作列，单击↑或者↓，调整安全组过滤的优先级。

优先级	安全组 ID	安全组名称	操作
1	sg-01313132685403	sg-01313132685403	↑ ↓ ⌕
2	sg-01313132685403	sg-01313132685403	↑ ↓ ⌕
3	sg-01313132685403	sg-01313132685403	↑ ↓ ⌕

6. 单击保存，完成修改。

调整出入站规则

1. 在安全组页签，可查看到实例当前所有的安全组。
2. 在安全组列表，单击安全组 ID 名称，跳转至 [安全组](#) 页面。
3. 找到需修改的安全组规则，在其操作列，单击编辑，可以重新编辑安全组规则。

导入安全组规则

1. 在 [安全组](#) 页面，选择需要的安全组，单击具体的安全组 ID/名称。

2. 在入站规则或者出站规则页签上，单击**导入规则**。
3. 在弹出的对话框中，选择已编辑好的入站/出站规则模板文件，单击**开始导入**。

说明：

- 如果需要导入规则的安全组下已存在安全组规则，建议您先导出现有规则，否则导入新规则时，将覆盖原有规则。
- 如果需要导入规则的安全组下没有安全组规则，建议您先下载模板，待编辑好模板文件后，再将文件导入。

克隆安全组

1. 在 [安全组](#) 页面，在列表的**操作列**选择**更多 > 克隆**。
2. 在弹出的对话框中，选定目标地域、目标项目后，单击**确定**。
若新安全组需关联 CVM，请重新进行管理安全组内云服务器。

删除安全组

1. 在 [安全组页](#)，选择需要删除的安全组，在**操作列**选择**更多 > 删除**。
2. 在弹出的对话框中，单击**确定**。
若当前安全组有关联的 CVM 则需要先解除安全组才能进行删除。

更多参考

关于安全组，更多的信息，请参见 [安全组概述](#)。

管理 Embedding 功能

最近更新时间：2024-05-13 15:53:51

腾讯云向量数据库（Tencent Cloud VectorDB）默认开通 Embedding 功能。您可以在控制台查看 Embedding 处理 Token 的速率、调整 Embedding 模型规格、关闭 Embedding 功能或者再次开通 Embedding 功能。

功能描述

默认开通 Embedding 功能，当前实例下所有集合均可使用 Embedding 功能。具体支持如下操作：

- 支持开通、关闭 Embedding 功能。
- 支持根据业务需求灵活调整 Embedding 规格。

模型规格

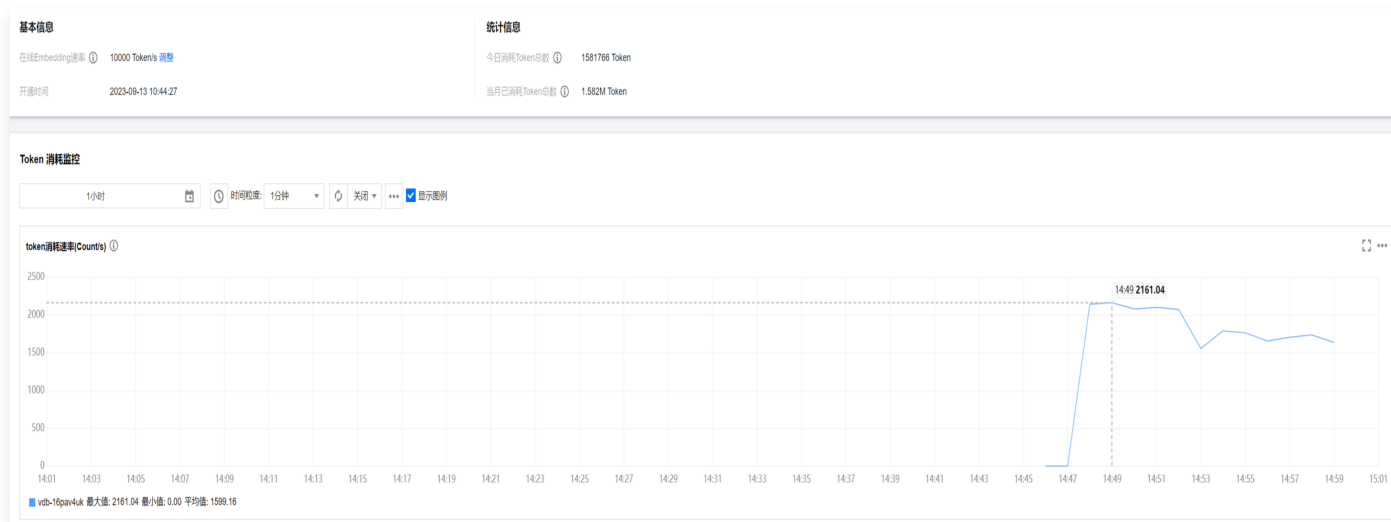
腾讯云向量数据库（Tencent Cloud VectorDB）Embedding 功能当前支持的速率为 10000 Token/s。如需更高规格，请[提交工单](#)申请。

说明：

在 Embedding 模型中，Token 是指文本数据处理的基本单元。通常在文本中，一个 Token 可以是一个字或词，也可以是一个标点符号。

查看 Embedding 监控数据

- 使用腾讯云账号登录 [向量数据库控制台](#)。
- 在实例列表中，找到需开通 Embedding 的实例，单击实例 ID 进入实例详情页面。
- 切换到 Embedding 页签，查看 Embedding 基本信息与监控数据。



界面区域	界面参数	参数含义
基本信息	在线 Embedding 速率	指 Embedding 每秒处理 Token 的速率。具体速率对应的规格层级，先参见 模型规格 。 单击调整，可调整层级。
	开通时间	Embedding 功能开通时间。
统计信息	今日消耗 Token 总数	统计当日从零点到当前处理的 Token 数量。
	当月已消耗 Token 总数	统计当月第一天零点到当前处理的 Token 数量。
Token 消耗监控	监控处理 Token 速率的变化趋势。 - 在时间框区域单击 ，选择所需监控数据的时间跨度。 - 在时间粒度后面的下拉列表选择监控粒度，即监控数据的采集粒度。 - 单击 ，可刷新监控数据，并更新监控视图。 - 单击关闭，可关闭数据更新；在关闭的下拉菜单，可设置更新数据的周期，即每隔多久更新数据。 - 在监控视图的右上角，单击 ，可全屏展示监控视图，单击 ，可将监控数据与视图导出。	

调整 Embedding 规格

1. 使用腾讯云账号登录 [向量数据库控制台](#)。
2. 在实例列表中，找到需开通 Embedding 的实例，单击实例 ID 进入实例详情页面。
3. 切换到 Embedding 页签，在基本信息区域，单击在线 Embedding 速率后面的调整。

基本信息		统计信息	
在线Embedding速率 ⓘ	2000 Token/s 调整	今日消耗Token总数 ⓘ	0 Token
开通时间	2023-09-11 16:01:39	当月已消耗Token总数 ⓘ	0 Token

4. 在调整在线 Embedding 规格窗口，可看到当前规格下每秒处理 Token 的速率，以及QPS 数据，单击提交工单，可申请更高规格。

调整在线 Embedding 规格

当前默认提供规格为10,000 Token/s，如需更高规格请[提交工单](#) [申请](#)

规格	限速 (Token/s)	QPS ⓘ
level 1	10000	100

费用: **0元** (Embedding 功能目前处于内测阶段，暂时免费)

确定

取消

关闭 Embedding 功能

如果确认不需要使用 Embedding 功能，可进行关闭。

⚠ **注意：**

- 关闭 Embedding 功能为**数据库实例级别**的操作。关闭后，已开启 Embedding 功能的集合，将无法再使用 Embedding 功能，**不可直接通过 `/document/upsert` 插入向量字段，仅可通过 `/document/update` 更新文本字段与向量字段。**
- 如果关闭 Embedding 功能时，有正在进行的 Embedding 任务还未结束，则该任务继续进行不会中断。

1. 在 Embedding 页签中，单击**关闭 Embedding 功能**。

关闭 Embedding 功能

基本信息

在线 Embedding 速率 ⓘ

1000 Token/s

开通时间

2024-03-20 15:40:39

统计信息

今日处理 Token 总数 ⓘ

0 Token

当月已处理 Token 总数 ⓘ

885 Token

2. 在**关闭 Embedding 功能**窗口，确认影响说明，单击**确定**。

开通 Embedding 功能

关闭 Embedding 功能之后， 您可以根据需要随时开通再次使用 Embedding 功能。

1. 在 Embedding 页签中，单击**开通 Embedding 功能**。

实例详情 实例监控 密钥管理 安全组 Embedding

Embedding功能

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。开启 Embedding功能并在创建 Collection时配置模型，在插入、更新和相似性检索数据时可直接传入原始文本，向量数据库会自动将原始文本进行转换，生成对应的向量数据后插入数据库或进行相似性计算，大幅提高业务接入效率。

开通Embedding功能

2. 在开通 Embedding 功能窗口，确认模型规格，单击**确定**。

相关操作

使用 Embedding 功能，需要在通过 API 建表时，设定 Embedding 模型，并配置相关参数，具体使用示例，请参见 [连接并写入原始文本](#)。

登录 DMC

最近更新时间：2024-05-13 17:15:12

腾讯云向量数据库（Tencent Cloud VectorDB）支持通过 DMC（Database Management Center，DMC）可视化管理数据库。文本介绍快速登录 DMC 个人版客户端的操作。

前提条件

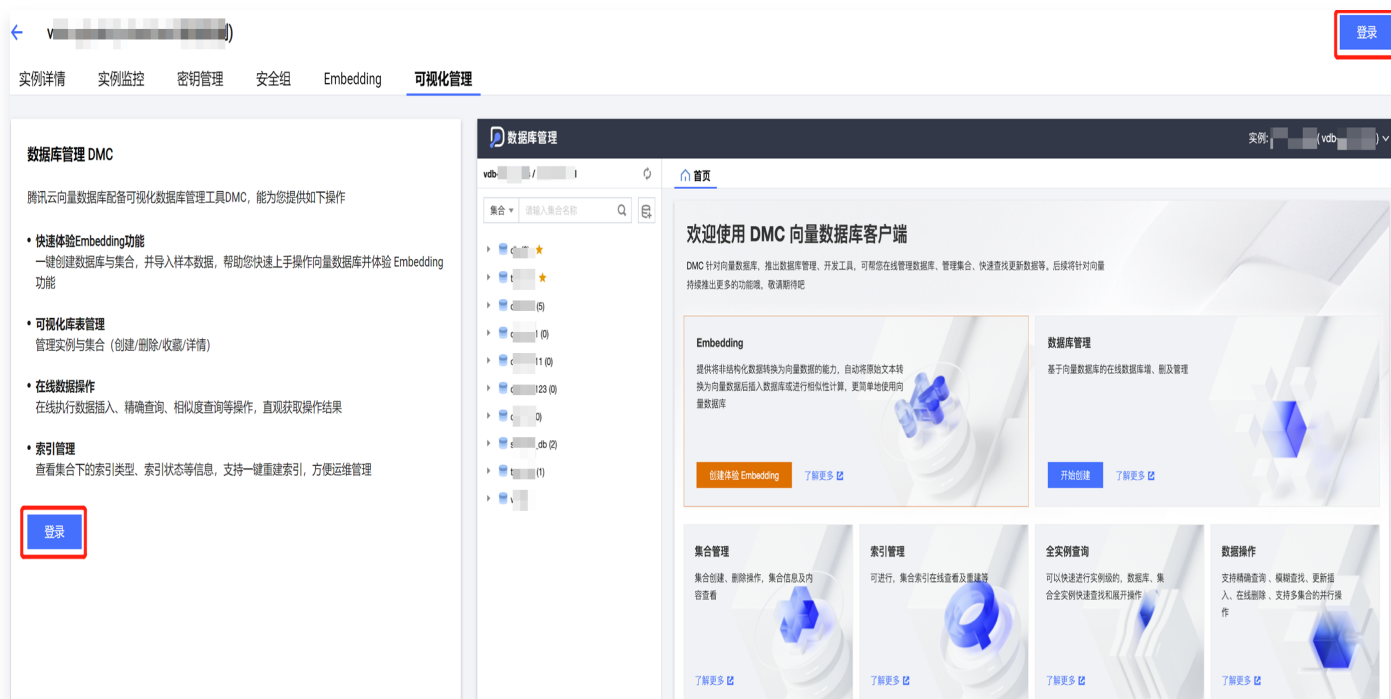
- 已 [创建数据库实例](#)。
- 实例状态为 **运行中**。

说明：

- 2023年09月16日 00:00 之前创建的向量数据库实例，需 [提交工单](#) 将内核版本升级为最新版本，才可使用 DMC。
- 2023年09月16日 00:00 及之后创建的向量数据库实例，可直接使用 DMC。

操作步骤

- 使用腾讯云账号登录 [向量数据库控制台](#)。
- 在实例列表中，找到需登录 DMC 的实例。
- 单击实例 ID 进入实例详情页面，切换到 **可视化管理** 页签。
- 单击右上角的**登录**，或左下方的**登录**，如下图所示。



5. 跳转至数据库管理 DMC 登录界面，如下图所示，输入账户信息及其对应密钥，单击**登录**。账户信息与密钥可在密钥管理页面获取，具体操作，请参见 [密钥管理](#)。



6. 进入向量数据库 DMC 首页，如下图所示。



7. 通过 DMC 管理数据库，具体操作及其指引文档，请参见下表。

数据库操作	操作明细	参见文档
Database 级操作	新建、删除数据库	数据库管理
Collection 级操作	新建、删除集合	集合管理
Document 级操作	插入、更新、精确查询、相似度检索文档	数据操作
查询操作	查询数据库、查询集合配置信息	查询数据库与集合
Index 管理	重建索引	索引管理

切换网络

最近更新时间：2024-12-25 11:51:22

腾讯云向量数据库依托于私有网络（Virtual Private Cloud，简称 VPC）技术，提供了一个安全、隔离的云上网络环境。VPC 技术允许在腾讯云中构建独立的网络空间，确保用户资源享有专属网络服务，同时实现不同 VPC 之间的逻辑隔离，保障网络的安全性和私密性。当购买数据库实例后，发现初始配置的私有网络不再能满足性能需求，或者需要对当前网络进行维护和升级，可以通过控制台操作将数据库实例迁移到其他私有网络环境。

使用须知

- 切换网络过程中，外网连接会引起闪断，请确保客户端程序具备重连机制。
- 切换网络之后，将导致实例内网 IP 发生变化，新的网络地址立即生效，原地址上所有连接将断开，请在业务低谷期谨慎执行。
- 切换之后，原网络地址保留时长最长为15天。超出释放时间后，旧的访问 IP 会失效，请及时修改客户端程序，进行业务切换。

操作步骤

1. 使用腾讯云账号登录 [向量数据库控制台](#)。
2. 在实例列表中，找到需切换网络的实例。
3. 单击目标实例 ID，或在其操作列，单击管理，进入实例详情页面。
4. 在网络信息区域，单击私有网络后面的更改网络。
5. 在修改网络小窗口，配置新的私有网络。
 - 在网络后面的下拉列表，分别选择新的私有网络和子网，也可以单击新建私有网络和新建子网，跳转至私有网络控制台页面，创建新的私有网络之后，再在网络后面的下拉框选择新创建的私有网络。
 - 在新IP地址后面的下拉列表，选择新的 IP 地址分配方式，支持自动分配与指定地址。若选择指定地址，则需在后面的输入框，根据选择的私有网络的CIDR（IP 地址范围）指定具体的 IP 地址。
 - 在旧 IP 地址后面的下拉列表，选择旧地址释放的时间，支持立即释放、1天后释放、2天后释放、3天后释放、7天后释放、15天后释放。

6. 单击**确定**，完成配置，等待切换任务执行完成，实例状态恢复为**运行中**，即可正常使用新的私有网络环境。

开启外网访问

最近更新时间：2025-04-14 19:38:32

腾讯云向量数据库（Tencent Cloud VectorDB）同时支持内网安全连接与公网访问能力。本文提供通过控制台快速启用公网访问权限的配置指南，并附注外网访问的安全策略建议，帮助开发者在保障数据安全的前提下灵活应用。

说明：

如果您的 Linux [云服务器 CVM](#) 与腾讯云向量数据库位于同一地域和 VPC 内，那么可以利用腾讯云内网实现快速连接。为确保网络互通，请确认 CVM 与向量数据库安全组的入站和出站规则。具体信息，可参见[安全组](#)。

使用须知

注意：

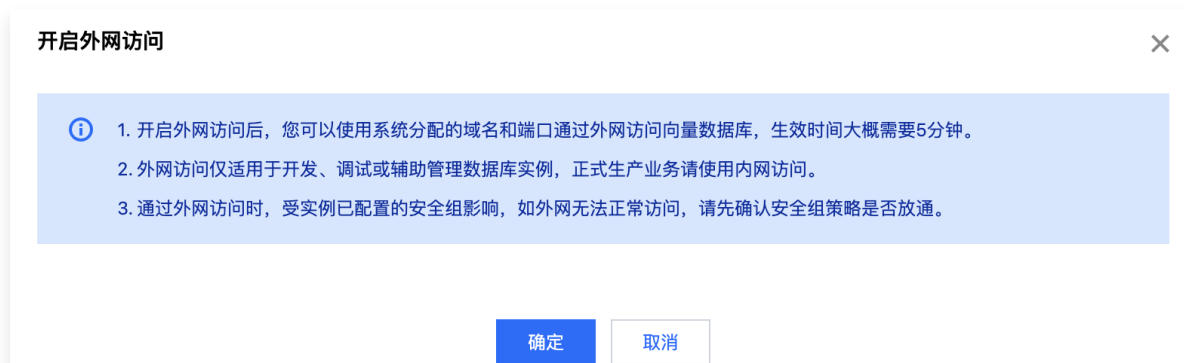
向量数据库对外网访问方式进行了规范化升级，将原本通过白名单访问外网的方式修改为实例安全组控制。

- 对于已经使用原始配置白名单列表开启外网访问的数据库实例，使用逻辑和白名单列表将保持不变。
- 但若关闭并重新开启外网访问，系统将自动切换至新的外网访问方式，建议提前记录白名单列表，并在重新开启外网访问后，在实例安全组中配置访问规则。

- 开启外网访问之后，系统自动分配域名与端口，生效时间约5分钟。
- 外网访问仅适用于开发、调试或辅助管理数据库实例，正式生产业务请使用内网访问。
- 通过外网访问时，受实例已配置的安全组影响，如外网无法正常访问，请先确认安全组策略是否放通。

操作步骤

- 使用腾讯云账号登录 [向量数据库控制台](#)。
- 在实例列表中，找到需开启外网访问的实例。
- 单击实例 ID 进入实例详情页面，在**网络信息**区域，单击**外网地址**后面的**开启**。
- 在**开启外网访问**的小窗口，阅读开启外网的使用说明，如下图所示。



5. 单击**确定**，等待任务执行完成，在**网络信息区域的外网地址**，可查看系统分配的外网地址。

- 单击，可复制外网地址。
- 单击**关闭**，可关闭外网功能。

⚠ 注意：

关闭后将无法通过外网地址访问数据库实例。请确认您的业务没有使用外网访问地址，以免带来不必要的损失。

- 单击**编辑白名单**，若外网域名为 `*.tencentcdb.com`，跳转至**安全组**页签，为实例绑定安全组进行网络安全控制；若外网域名为：`*.tencentclb.com`，编辑访问实例的白名单列表。

网络信息

私有网络 
内网 IP 
端口 80 
访问地址 [http://!\[\]\(240709b4bf488c5389abc96fa5be38e1_img.jpg\) 连接并写入数据库](#)
外网地址  [关闭](#) [编辑白名单](#)

6. 在本地使用外网地址连接向量数据库。使用 HTTP 连接数据库的示例，如下图所示。

```
[root@localhost ~]# curl -i -X GET \
> -H 'Content-Type: application/json' \
> -H 'Authorization: Bearer account=root&api_key= \
> http:// database/list
HTTP/1.1 200 OK
Content-Length: 57
Content-Type: application/json
{"code":0,"msg":"operation success","databases":["test"]}
```

备份与回档

备份概述

最近更新时间：2025-04-23 16:55:01

腾讯云向量数据库提供数据备份与回档服务，支持预设备份时间点并周期性执行主动备份，同时允许随时手动触发即时备份操作；所有备份均以高可靠快照形式存储，通过实例克隆功能可实现数据回档，恢复业务状态。

说明：

V2.4 版本起，支持产品化的备份克隆功能，若需要使用该功能，请 [提交工单](#) 升级实例版本。

备份方式

自动备份

由系统按照预设的备份策略（例如备份周期、备份时间窗）定时触发备份任务的备份方式，主要用于日常业务连续性保障，以防范故障、数据误删等常规风险。

手动备份

用户在控制台主动发起立即执行的备份方式，适用于重大系统升级、架构变更之前的数据快照，或应对突发性风险手动应急保护。

数据回档

克隆实例：腾讯云向量数据库支持在同地域，基于备份集克隆一个完整的新实例，通过克隆操作将所选备份集对应时间点的备份文件恢复到一个新的实例中，该实例与源环境完全解耦，且实例的数据和备份文件一致，您可以使用克隆功能来分析以往的数据，也可以通过修改访问地址的方式，交换克隆的新实例和原有实例的 IP 来达到回档的目的。

注意：

- 当前仅支持回档到指定备份集对应的时间点，暂时无法回档到任意时间点。因此，若业务在进行敏感变更之前，可以发起一次手动备份。
- 当前仅支持通过克隆实例的方式回档到新实例，暂不支持回档到当前实例。
- 当前仅支持整实例克隆，暂不支持库表级别克隆。
- 若使用 AI 套件，则克隆后无法还原对应时间点的图片信息。

备份类型

快照备份：一种在存储层对云硬盘创建快照，捕获特定时间点的磁盘数据状态的数据备份方式。备份速度快，对数据库性能的影响小，且无需停机，可以轻松地从快照中恢复数据。

说明：

数据备份保留时长默认为7天，暂不允许修改，即用户可以回档至7天内全量备份集对应的时间点，到期后系统自动执行删除操作。

查看备份记录

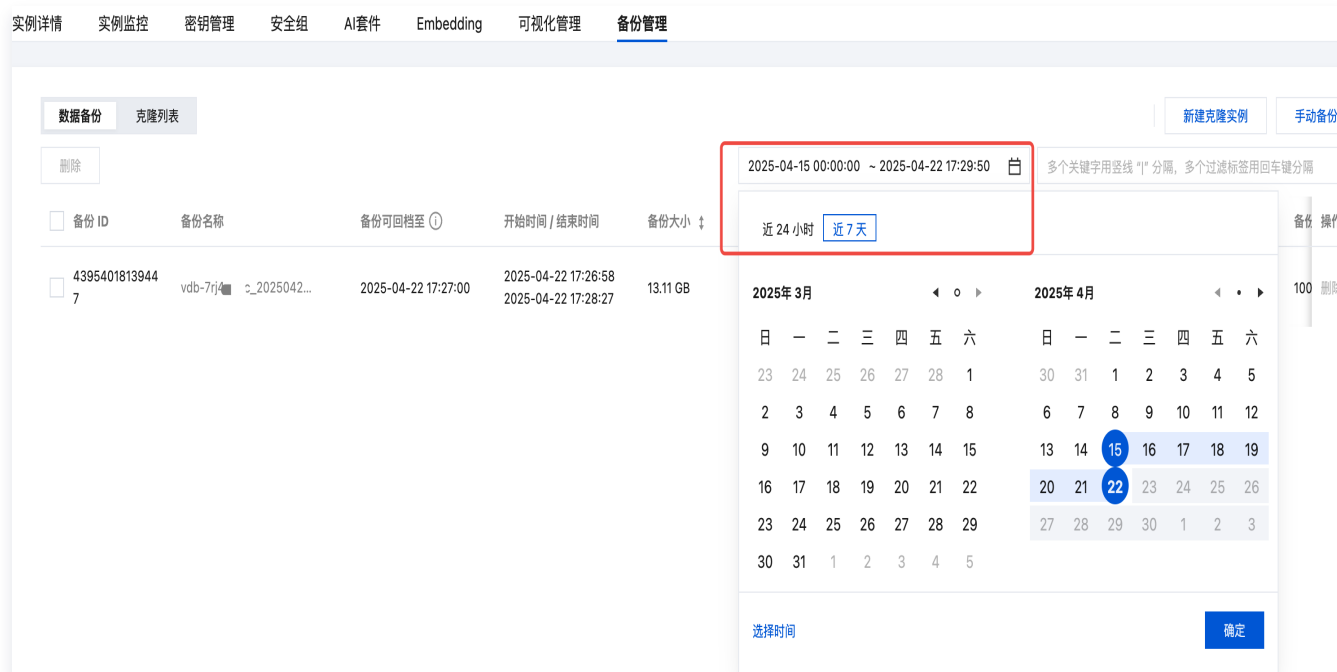
最近更新时间：2025-04-22 22:06:21

操作场景

数据备份记录了每一次自动备份或手动备份的执行情况，为数据的全生命周期管理构建了可验证、可追溯的保障体系。在数据恢复前，可依据备份记录验证目标时间点备份的完整性和有效性，从而规避恢复失败的风险；当数据出现异常时，可通过备份生成时间、存储状态等元数据快速定位问题根源。例如，当业务数据因误删需紧急回滚时，用户可快速检索备份记录，确认最近有效快照的存储状态与时间戳，进而通过克隆实例验证数据一致性后执行精准恢复，确保业务连续性。

操作步骤

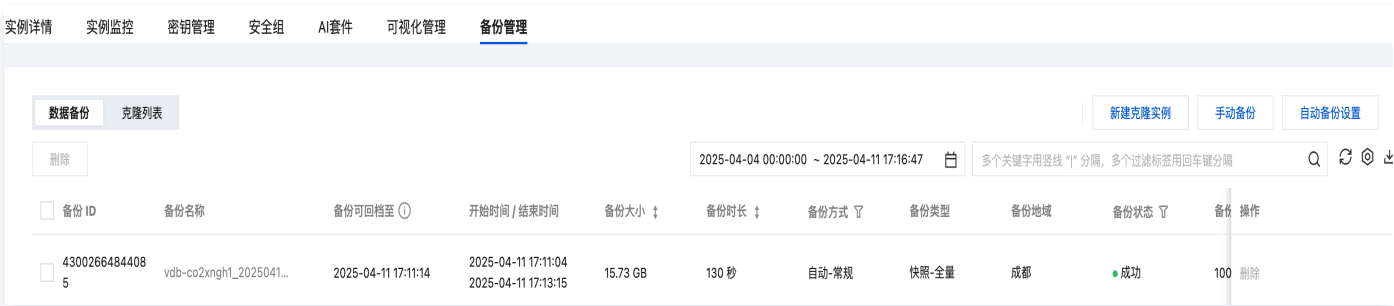
1. 使用腾讯云账号登录 [向量数据库控制台](#)。
2. 在实例列表中，找到需查看的实例。
3. 单击目标实例 ID，或在其操作列，单击**管理**，进入实例详情页面。
4. 切换至**备份管理**页签，可查看到所有的数据备份记录，可根据如下方式搜索备份数据。
 - 在时间框的下拉表中，可选择所需查看备份数据的时间范围，搜索备份数据。并且，您可直接选择查看近 24 小时，近7天的备份数据。如下图所示。



- 在搜索框的下拉列表，可根据多个关键字或多个过滤标签（备份 ID、备份名称、备份方式、备份状态）快速搜索备份数据，如下图所示。



5. 备份数据以列表的形式展示，如下图所示。在备份记录列表右上角，单击 ，可自定义备份列表显示的字段；单击 ，可导出当前页面的所有备份记录。



参数名	参数解释
备份 ID	备份任务的唯一标识符，用于区分不同的备份任务。将鼠标放在备份 ID，单击  ，可复制备份 ID。
备份名称	指备份任务的名称。 - 修改备份名称：将鼠标放在名称上，可单击 ，重新编辑名称。 - 名称要求：仅支持中英文、数字，特殊字符-_.(/)[]+=:;@，最长128 个字符。
备份可回档至	克隆实例时，可回档到该时间点的备份数据。
开始时间/结束时间	备份任务执行开始与结束的时间。
备份大小	备份数据大小，指自上一次备份后云盘的增量变化量，该增量按 1MB 为单位计算。
备份时长	备份任务完成所需的时长。
备份方式	备份任务的执行方式。 - 自动-常规：自动备份，根据用户预设的备份时间周期性自动发起备份任务。 - 手动：手动备份，由用户主动发起立即执行的备份任务。

备份类型	备份数据的具体类型，当前仅支持快照备份（Snapshot），捕获特定时间点的磁盘数据状态的数据备份方式。
备份地域	备份数据存储的地域。
备份状态	备份任务的当前状态。
备份进度	备份任务执行的进度。
过期时间	备份数据过期的时间点。备份数据将保留7天，到期后自动执行删除操作。
备份操作	对备份任务的操作选项。单击删除，可删除不再需要的手动备份记录。

自动备份

最近更新时间：2025-04-22 22:06:21

操作场景

腾讯云向量数据库的自动备份功能通过预设策略（如每日凌晨业务低峰期）自动触发全量快照生成，全程无需人工干预，有效应对硬件故障、误删数据或软件异常等风险场景。

操作步骤

说明：
在实例创建流程中，系统将自动执行首次全量备份。实例初始化完成后，用户可在备份管理页面查看对应的初始备份任务记录。

1. 使用腾讯云账号登录 [向量数据库控制台](#)。
2. 在实例列表中，找到需设置自动备份规则的实例。
3. 单击目标实例 ID，或在其操作列，单击**管理**，进入**实例详情**页面。
4. 切换至**备份管理**页面，单击**自动备份设置**。

实例详情

实例监控

密钥管理

安全组

AI套件

可视化

备份管理

数据备份

克隆列表

删除

2025-04-04 00:00:00 ~ 2025-04-11 17:16:47

多个关键字用竖线 "|" 分隔，多个过滤标签用回车键分隔

新建克隆实例

手动备份

自动备份设置

☐	备份 ID	备份名称	备份可回档至 ①	开始时间 / 结束时间	备份大小 ↓	备份时长 ↓	备份方式 ▾	备份类型	备份地域	备份状态 ▾	备份 操作
☐	43002664844085	vdb-co2xnght_2025041...	2025-04-11 17:11:14	2025-04-11 17:11:04 2025-04-11 17:13:15	15.73 GB	130 秒	自动-常规	快照-全量	成都	成功	100 删除

5. 在**自动备份设置**的弹窗，可在**备份开始时间**的下拉列表，选择自动备份任务可执行的时间范围。

- 说明：**
- 备份开始时间：默认每天00:00-04:00内发起自动备份任务，执行一次全量数据备份。
 - 数据备份保留时长：默认为7天，暂不允许修改，即用户可以回档至7天内全量备份集对应的时间点，到期后系统自动执行删除操作。
 - 备份周期：默认勾选周一至周日，每天定时备份，不允许修改。

自动备份设置

×

数据备份

常规设置

备份开始时间

00:00 - 04:00

▼

备份存储介质 *

SNAPSHOT

▼

数据备份保留时长

—

7

+

天

1. 默认保留7天，到期后备份集自动清理

2. 无法恢复到保留时段外的时间

备份周期 *

☒ 周一

☒ 周二

☒ 周三

☒ 周四

☒ 周五

☒ 周六

☒ 周日

预计下次备份时间

2025-04-12 00:00 ~04:00 ⓘ

确定

取消

6. 单击**确定**，完成设置。

手动备份

最近更新时间：2025-04-22 22:06:21

操作场景

腾讯云向量数据库的手动备份功能主要服务于需即时保障关键数据状态的场景，弥补自动备份周期间隙的数据保护盲区。例如：当监测到异常攻击或突发系统告警时，立即创建临时备份以锁定当前数据完整性，为关键数据提供即时保障，确保数据安全与可恢复性。

操作步骤

⚠ 注意：

一个实例每天最多可发起5次手动备份。

1. 使用腾讯云账号登录 [向量数据库控制台](#)。
2. 在实例列表中，找到需进行手动备份的实例。
3. 单击目标实例 ID，或在其操作列，单击管理，进入实例详情页面。
4. 切换至备份管理页面，单击手动备份。
5. 在手动备份弹窗，根据提示要求，设置备份名称。

手动备份

① 点击确认后立即发起备份任务，单个实例每天最多进行5次手动备份

备份对象 *

☒ 实例备份

备份名称

请输入备份名称

仅支持中英文、数字，特殊字符-_/()[]+=;:@，最长128 个字符

保留时间策略 *

☒ 自定义策略(即跟随自动备份保留时间策略)

确定

取消

6. 单击确定，完成设置。在备份列表中，可看到已生成一条备份记录，备份状态为备份中，等待任务执行完成即可。

克隆实例

最近更新时间：2025-04-23 16:55:01

操作场景

在当前实例数据出现严重问题，需要回滚到之前备份的状态时，您可以基于当前实例的备份集直接克隆一个新实例来快速恢复数据。克隆实例的数据和备份时间点保持一致，您可以使用克隆实例来分析历史数据，也可以修改访问地址的方式，交换克隆的新实例和源实例的 IP 地址来达到回档的目的，从而提高恢复数据的效率和准确性。

使用说明

⚠ 注意：

- 当前仅支持回档到指定备份集对应的时间点，暂时无法回档到任意时间点。因此，若业务在进行敏感变更之前，可以发起一次手动备份。
- 当前仅支持通过克隆实例的方式回档到新实例，暂不支持回档到当前实例。
- 当前仅支持整实例克隆，暂不支持库表级别克隆。
- 若使用 AI 套件，则克隆后无法还原对应时间点的图片信息。

操作步骤

1. 使用腾讯云账号登录 [向量数据库控制台](#)。
2. 在实例列表中，找到需克隆的实例。
3. 单击目标实例 ID，或在其操作列，单击**管理**，进入**实例详情**页面。
4. 切换至**备份管理**页面，单击**新建克隆实例**。
5. 在**向量数据库 VectorDB 克隆实例**购买页，根据下表的参数解释，设置克隆实例的备份集与规格。

向量数据库 VectorDB 克隆实例

[产品价格](#) [产品控制台](#)

基础配置

计费模式

包年包月

按量计费

地域

深圳金融

处于不同地域的云产品内网不通，创建成功后**不支持切换地域**；建议选择最靠近您客户的地域，可降低访问时延。

克隆配置

恢复方式

按照备份集

备份集

2025-04-14 16:48:31

源实例信息					
实例 ID	vdb-nqcy704c	实例名称	vdb-nqcy704c	所属地域	深圳金融
所属网络	深圳金融 - AutoName_20230421_...	类型与部署模式	高可用版 三可用区	实例规格	计算型 3 节点 * (1 核 2G 20G 磁盘)

规格信息

实例类型

高可用版

一主多从的分布式架构，支持配置多个数据副本，性能强劲

部署模式

三可用区

支持配置2个以上的数据副本，提供更高的可靠性和可用性

节点数量

3

支持选择3 - 30节点

节点规格

节点类型

计算型

计算型适用于QPS高、延迟敏感的场景，更侧重于向量数据的快速查找和检索

已选择

计算型 1 核 2 GB

实例规格选型指导 2

节点类型	CPU (核)	内存 (GB)	建议向量规模 (1536维度) ①	参考价格
☐ 计算型	1	2	250000	
☐ 计算型	2	4	500000	
☐ 计算型	4	8	1000000	
☐ 计算型	8	16	2000000	
☐ 计算型	16	32	4000000	
☐ 计算型	24	48	6000000	
☒ 计算型	32	64	8000000	

节点磁盘类型

云硬盘

节点硬盘容量

20 GB

500 GB

1000 GB

20

GB

(步长为: 10GB)

节点磁盘容量取值范围: 20 - 1000 GB

网络和安全组

网络

深圳金融

AutoName_20230421_112635

CIDR: 10.0.2.0/24, 子网内可用IP: 253个/250个

当前网络选择下, 仅“深圳金融”网络的主机可访问数据库 [新建私有网络](#) [新建子网](#)

安全组

请选择

如您业务需要放通其他端口, 您可以 [自定义安全组](#)

实例设置

标签

标签键

标签值

删除

添加

键值粘贴板

如现有的标签不合适, 您可以去控制台 [新建标签键/标签值](#)

实例名

创建后设置

立即设置

请输入实例名称

实例名称由中英文、数字、下划线组成, 最长60个字符; 批量购买实例时, 会在自定义实例名称的尾部添加数字序号

自动续费

☐ 账户余额足够时, 设备到期后按月自动续费

服务条款

☐ 我已阅读并同意 [《云数据库服务条款》](#)、[《服务等级协议》](#)

参数类别	参数	参数解释
基础配置	计费模式	选择克隆实例的计费方式。购买之后, 支持按量计费转包年包月计费。 - 包年包月: 适用于业务规划明确, 长期稳定的场景。 - 按量计费: 适用于业务发展有较大波动性, 无法进行准确预测的场景。
	地域	默认与源实例备份集时间点的实例地域保持一致, 不可修改。
克隆配置	恢复方式	仅支持按照备份集恢复数据, 指通过预先生成的备份文件 (即备份集), 将数据库状态还原到备份创建时的时间点。
	备份集	在下拉列表中, 选择备份集的时间点, 克隆实例便基于该时间点的备份集恢复数据。
	源实例信息	备份集所对应的源实例的基础信息, 包含: ID、名称、地域、网络、类型以及规格。 说明: 备份集若生成在实例规格变更之前, 则显示变更之前的规格信息。
规格信息	实例类型	默认与源实例备份集时间点的实例类型一致, 不可修改。

	部署模式	默认与源实例备份集时间点的实例部署模式一致，不可修改。
	节点数量	默认与源实例备份集时间点的实例节点数量一致，不可修改。
	节点规格	- 节点类型：默认与源实例备份集时间点的实例节点类型一致，不可修改。 - 节点规格：为克隆实例选择与源实例相同或更高的配置规格。
	节点磁盘类型	固定为云硬盘。
	节点硬盘类型	在滑轴上选择所需的磁盘容量，支持选择与源实例相同或更大的磁盘规格。取值范围：[源实例的硬盘容量,1000]，单位：GB。
网络和安全	网络	在下拉列表分别选择已配置的私有网络及子网。若现有网络不满足需求，请单击 新建私有网络 或 新建子网 ，创建所需的网络环境。具体操作，请参见 创建私有网络 。
	安全组	默认与源实例的安全组一致，可在下拉列表中重新为克隆实例设置安全组。单击 自定义安全组 ，也可以设置新的安全组。具体操作，请参见 安全组 。
实例设置	标签	为克隆实例设置标签。 - 在下拉列表选择已配置的标签。 - 若现有标签不合适，请单击添加，可以添加标签键与标签值。具体操作，请参见 标签管理。
	实例名	在输入框设置实例名称。仅支持长度不超过60的中文/英文/数字/-/_。
	自动续费	计费模式为包年包月，需配置是否需设置自动续费，即实例到期时，从云账户直接扣除一个月的费用。实例购买之后，支持取消自动续费。具体操作，请参见 续费（包年包月） 。
	服务条款	勾选 我已阅读并同意云数据库服务条款 。
	时长	计费模式为包年包月，在下拉列表，按整月或整年选择购买实例时长，最大支持购买时长为3年。
	数量	[1,10]

6. 在**配置费用**，确认计费金额，单击**立即购买**。在实例列表，可看到新克隆的实例状态为**从源实例克隆中**，等待状态更新为**运行中**，即可使用克隆实例。