

知识湖存储

用户指南

文档版本

01

发布日期

2025-09-29



版权所有 © 华为云计算技术有限公司 2025。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为云计算技术有限公司

地址：贵州省贵安新区黔中大道交兴功路华为云数据中心 邮编：550029

网址：<https://www.huaweicloud.com/>

目 录

1 产品介绍.....	1
1.1 什么是知识湖存储.....	1
1.2 产品优势.....	2
1.3 应用场景.....	2
1.4 产品功能.....	2
1.5 计费说明.....	3
2 申请公测.....	4
3 知识仓实例管理.....	5
3.1 知识仓实例介绍.....	5
3.2 管理知识仓实例.....	5
4 VPCEP 终端节点对接 LMS 知识仓实例.....	9
5 查看监控指标.....	17
5.1 支持的监控指标.....	17

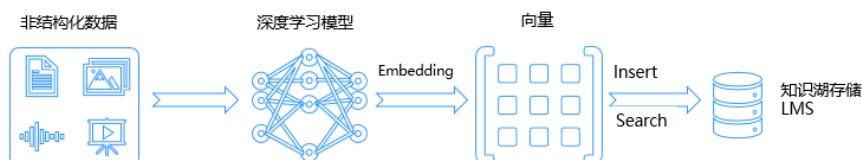
1 产品介绍

1.1 什么是知识湖存储

知识湖存储（Long-term Memory Service，LMS），是一款用于存储、检索、分析多维向量数据和标量数据的云原生服务。

知识湖存储服务支持多种索引类型和相似度计算方法，支持千亿向量规模和毫秒级查询时延。

知识湖存储不仅能为大模型提供外部知识库，提高大模型回答的准确性，还可广泛应用于推荐系统，自然语言处理，智能客服等AI领域。



访问方式

提供了Web化的服务管理平台，即管理控制台和基于HTTPS请求的API（Application programming interface）和SDK的管理方式。

- 控制台方式
通过控制台可以创建知识仓实例。
- API方式
创建知识仓实例后，可以通过Restful API来创建或删除collection、插入向量、查询向量等操作。
- SDK方式
创建知识仓实例后，可以通过python SDK来创建或删除collection、插入向量、查询向量等操作。

1.2 产品优势

知识湖存储作为一种存储和检索向量数据的服务提供给用户，在高性能、高可用、大规模、简单易用、安全可靠等方面体现出显著优势。

- **高性能**：知识湖存储支持千亿级向量数据规模及毫秒级查询延迟。
- **高可用**：知识湖存储提供多副本高可用特性，其多可用区和分布式多节点的架构，可用性可达99.99%，显著提高系统的可靠性和容错性，确保知识湖存储服务在面临节点故障和负载变化等挑战时仍能正常运行。
- **大规模**：知识湖存储是云原生的服务，架构支持水平扩展，可支持百亿向量规模，满足 AI 场景下的向量存储与检索需求。
- **简单易用**：支持丰富的向量检索能力。用户通过 HTTP API 或者 SDK 接口即可快速使用知识湖存储服务，开发效率高。
- **安全可靠**：充分复用华为云对象存储的安全可信能力，实现安全隔离和全链路数据安全可信保障。

1.3 应用场景

知识湖存储可进行高性能向量存储和检索，主要适用于以下应用场景：

- **文本/图像检索**：向量数据库对输入的图像和文本信息进行向量存储与检索，会找到最匹配输入信息的文本或图像结果。
- **大规模知识库**：企业的私域数据存储在向量数据库中可构建外部知识库，帮助企业更好地管理和利用自己的数据资源。
- **问答系统**：向量数据库会基于问题信息进行向量存储与检索，并返回最相关的问题与对应的答案。
- **推荐系统**：向量数据库会基于用户特征进行向量存储与检索，最终筛选用户可能感兴趣的物品推荐给用户。

1.4 产品功能

表 1-1 知识湖存储功能

功能名称	功能描述
向量和标量存储	将数据以向量和标量的形式存储在对象存储服务OBS中
向量检索	计算目标向量与OBS中向量之间的相似度，获取TOP K个最相似的向量
标量检索	基于数据的标量字段进行精确匹配的检索
混合检索	稠密向量 Dense vector 和 稀疏向量 Sparse vector 混合检索

1.5 计费说明

在使用知识湖存储中的知识检索功能时，需要先创建知识仓实例，使用知识仓实例，会收取一定的费用。

计费说明

资费项	计费项	含义	适用的计费模式	计费公式
知识检索 计费	知识仓实例 费用	根据知识仓实例的CU规格、CU数量以及使用时长（小时）进行按需计费。 使用时长不足一小时的，以实际使用时长为准。	按需计费	CU规格单价 * CU数量 * 使用时长（小时）

计费示例

须知

以下示例中出现的费用价格仅供参考，实际价格请以控制台显示为准。

假设某用户于2025年5月27日14:00:00创建了一个知识仓实例，该知识仓实例的CU规格为2xlarge，CU数量为3个，假设2xlarge的单价为0.3元/个/小时，那么截止2025年5月27日18:36:00，该知识仓实例的计费如下：

2xlarge规格的计费单价 * CU数量 * 使用时长 = 0.3元/个/小时 * 3个 * 4.6小时 = 4.14元

说明

按需计费模式，使用时长不足一小时的根据实际使用时长计费，而36分钟换算为小时，为0.6小时，因此本示例中实际使用时长为4.6小时。

2 申请公测

知识湖存储服务 LMS目前正在公测中，支持用户申请使用。

操作步骤

步骤1 登录[管理控制台](#)。

步骤2 在系统首页，单击“服务列表”中的“存储 > 对象存储服务 OBS > 知识湖存储 > 知识检索”。

步骤3 单击“申请公测”，进入申请公测页面。

步骤4 在公测申请页面，请根据实际情况填写企业规模、研发人员比例、应用场景、业务当前阶段、业务描述等申请信息。

步骤5 勾选“同意《公测试用服务协议》”，单击“申请公测”。

----结束

公测申请提交后，5个工作日内审核结果将发送到您的邮箱和手机。

3 知识仓实例管理

3.1 知识仓实例介绍

知识仓，简称仓，是LMS中用来组织、存储和管理数据的单位。

LMS自顶向下按照仓、集合、向量的结构存储数据。

仓中包含至少1个或多个集合，集合中包含至少1个或多个向量，集合中也可以包含标量。集合中的向量用于相似性查询。

3.2 管理知识仓实例

创建知识仓实例

单个账号支持创建的知识仓实例数量不超过16个。

步骤1 进入[LMS控制台](#)，单击“创建知识仓实例”。

步骤2 如[图3-1](#)所示，根据界面提示配置参数，参数说明如[表3-1](#)所示。

图 3-1 创建知识仓实例界面

创建知识仓实例

基础配置

计费模式

按需计费

区域

可用区

可用区1可用区2可用区3

实例配置

知识仓实例名称

请输入实例名称

CU规格

请选择

CU数量

-1+

表 3-1 参数说明

参数	是否 必选	说明	示例
计费模式	是	当前只支持按需计费模式。	按需 计费

参数	是否 必选	说明	示例
区域	是	租户所在的区域，当前区域请在界面左上方选择。 选择和云服务器同一个区域。	华北-北京四 中国-香港 EU-Dublin
可用区	是	可用区指在同一区域下，电力、网络隔离的物理区域，可用区之间内网互通，不同可用区之间物理隔离。 建议您选择和云服务器相同的可用区。	可用区1
知识仓实例名称	是	知识仓实例名称，全域唯一。 知识仓的命名规范为：{用户自定义前缀}-{当前控制台所在区域的区域名}-{账号ID}。 需要满足如下规范： • 输入长度不能小于3个字符。 • 输入长度不能超过63个字符。 • 支持小写字母、数字、中划线（-）。 • 中划线（-）不得出现在字符串的头部或者尾部。 • 只能以小写字母开头。	storeprefix-cn-north-4-domainid
CU规格	是	CU即计算单元（Compute Unit），用于衡量知识仓实例计算能力的指标，在数据检索和分析时用于执行计算任务。 可选择的知识仓实例计算单元的规格。根据业务规模选择的对应的规格。 CU规格取值范围：xlarge, 2xlarge, 4xlarge, 8xlarge, 16xlarge。	4xlarge
CU数量	是	知识仓实例配置的CU数量，单位：个。 该数据会占用CU资源配额数量。 CU规格与CU配额的换算关系是：xlarge占1个CU配额，2xlarge占用2个CU配额，以此类推，16xlarge占用16个CU配额。当CU配额不够时，请提工单申请。	1

步骤3 配置完成后，单击“下一步”。

步骤4 核对知识仓实例信息，确认无误后单击“立即创建”。

----结束

列举知识仓实例

- 步骤1** 进入[LMS控制台](#)。
- 步骤2** 在知识仓实例列表中查看所有知识仓实例。
- 结束

搜索知识仓实例

- 步骤1** 进入[LMS控制台](#)。
- 步骤2** 在知识仓实例列表上方的搜索框中单击左键，在一级下拉框中选择“名称”、“区域”、“状态”、“规格”、“创建时间”，然后在二级下拉框中选择您需要的选项，也可以输入关键字后再进行选择，搜索到的知识仓实例会展示在列表中。
- 或在知识仓实例列表上方的搜索框中直接输入关键字，按Enter键。搜索到的知识仓实例“名称”、“区域”、“状态”、“规格”、“创建时间”中包含关键字的所有知识仓实例会展示在列表中。
- 结束

删除知识仓实例

- 步骤1** 进入[LMS控制台](#)。
- 步骤2** 单击待删除的知识仓实例操作列的“删除”。
- 步骤3** 在“删除实例”弹窗，一键输入“DELETE”，然后单击“确定”。
- 结束

扩容知识仓实例

- 步骤1** 进入[LMS控制台](#)。
- 步骤2** 单击待扩容的知识仓实例操作列的“扩容”。
- 步骤3** 如[图3-1](#)所示，根据界面提示配置参数，参数说明如[表3-1](#)所示。
- 步骤4** 如所示，根据界面提示配置参数，参数说明如[表3-1](#)所示。
- 步骤5** 配置完成后，单击“确认”。
- 结束

4 VPCEP 终端节点对接 LMS 知识仓实例

华为云租户通过API、SDK方式访问LMS服务时，需要通过VPC终端节点对接LMS。本章节指导用户如何通过VPC终端节点（下文简称：VPCEP）对接LMS知识仓实例。

VPCEP对接LMS主要分为以下三大步骤：

1. [购买连接“接口”型终端节点服务的终端节点](#)
2. [创建内网域名](#)
3. [添加A类型记录集](#)

须知

用户购买VPC终端节点，创建内网域名，内网域名即用户使用API、SDK访问LMS过程中使用到的终端节点。

注意事项

- 当前LMS只提供同区域访问，只能在LMS提供服务的区域购买终端节点。
- 为保障业务冗余，建议购买多个终端节点对接LMS提供的不同服务名称。

购买连接“接口”型终端节点服务的终端节点

步骤1 登录[VPC终端节点控制台](#)。

步骤2 在“终端节点”页面，单击“购买终端节点”，进入“购买终端节点”页面。

在“购买终端节点”页面，根据提示配置参数购买连接“接口”型终端节点服务的终端节点。

更多参数详情请参见[购买连接“接口”型终端节点服务的终端节点](#)。

图 4-1 购买终端节点（云服务-接口型）

购买终端节点

基础配置

区域

西南-贵阳一

不同区域的云服务产品之间内网互不相通；请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。

计费模式

按需计费

服务类别

云服务

按名称查找服务

服务名称

cn-

验证

已找到服务 服务类型：接口

创建内网域名

终端节点类型

接口终端节点

实例类型

专业型

基础型

了解实例详情

网络类型

IPv4

双栈

虚拟私有云

vpc-

创建虚拟私有云

查看虚拟私有云

子网

subnet-

创建子网

查看已有子网

可用IP数：249

表 4-1 终端节点配置参数

参数	说明
区域	终端节点所在区域。不同区域的资源之间内网不互通。请选择靠近您的区域，可以降低网络时延、提高访问速度。
计费方式	仅支持按需计费。按终端节点的实际使用时长计费，可以随时开通/删除终端节点。

参数	说明
服务类别	可选择“云服务”或“按名称查找服务”。 - 云服务：当您要连接的终端节点服务为云服务时，需要选择“云服务”。 - 按名称查找服务：当您要连接的终端节点服务为用户私有服务时，需要选择“按名称查找服务”。
服务名称	LMS只支持“服务类别”选择“按名称查找服务”。 从“查询知识仓实例页面”中找到“私网连接”的名称，然后在这里按照名称进行查找。 注意 如果输入服务名称后，查找不到任何信息，需要把账号ID添加到终端节点服务的白名单中。
创建内网域名	如果您想要以域名的方式访问终端节点，则选择“创建内网域名”，终端节点创建完成后，即可通过内网域名直接访问终端节点。
虚拟私有云	选择终端节点所属的虚拟私有云。
子网	选择终端节点所属的子网。
IPv4地址	终端节点的私网IP。可选择“自动分配IPv4地址”或“手动指定IP地址”。
访问控制	用于设置允许访问终端节点的IP地址或网段。 - 开启：只允许白名单列表中的IP地址或网段访问终端节点。 - 关闭：允许任何IP地址或网段访问终端节点。
白名单	用于设置允许访问的IP地址或网段，最多支持添加20个记录。 请输入允许访问的IP地址或网段，不支持格式：0.0.0.0和x.x.x.x/0。
标签	可选参数。 终端节点的标识，包括键和值。可以为终端节点创建10个标签。

步骤3 参数配置完成，单击“立即购买”，进行规格确认。

- 规格确认无误，单击“提交”，任务提交成功。
- 参数信息配置有误，需要修改，单击“上一步”，修改参数，然后单击“提交”。

步骤4 购买成功后，在终端节点控制中，找到创建出来的vpc终端节点，单击ID，进入到该终端节点“基本信息”页中。

图 4-2 vpc 终端节点基本信息

基本信息

访问控制

标签

ID

虚拟私有云

付费方

终端节点服务名称

创建时间

访问控制

描述

id=

vpce-

服务使用方

cn-

2025/05/09 11:23:51 GMT+08:00

--

状态

类型

实例类型

IPv4地址

IPv6地址

内网域名

已接受

接口

专业型

192.168.8.1

--

⚠ 注意

立即购买后，状态为：“待审批”，请联系客服进行审批；待审批后，终端节点才会真正创建成功，此时状态为：“成功”。

----结束

创建内网域名

步骤1 登录DNS控制台。

步骤2 在左侧导航栏选择“内网域名”，进入“内网域名”页面。

步骤3 在页面右上角，单击“创建内网域名”。

步骤4 在“创建内网域名”页面中，输入域名及相关参数。

更多参数详情请参见[创建内网域名](#)。

图 4-3 创建内网域名

创建内网域名

* 域名 ?

dwr-lms.ci-2.myhuaweicloud.com

例如：example.com

☐ 子域名递归解析代理 ?

* 区域 ?

西南-贵阳一

* VPC ?

vpc-

查看VPC

内网域名关联的VPC，只有此VPC中的ECS服务器才能解析创建的内网域名。

* 企业项目 ?

--请选择--

新建企业项目

标签

如果您需要使用同一标签标识多种云资源，即所有服务均可在标签输入框下拉选择同一标签，建议在TMS中创建预定义标签。查看预定义标签

在下方键/值输入框输入内容后单击添加，即可将标签加入此处

请输入标签键

请输入标签值

添加

您还可以添加20个标签。

描述

0/255

取消

确定

表 4-2 创建内网域名参数说明

参数	参数说明
域名	创建的内网域名名称。 输入"dwr-lms.{region}.myhuaweicloud.com"。
区域	内网域名所在区域。
VPC	内网域名要关联的VPC。 说明 关联的vpc需要与服务器（例如ECS）所在vpc保持一致，否则会导致内网域名解析不成功。
企业项目	内网域名关联的企业项目，用于将内网域名按照企业项目进行管理。 说明 仅当用户使用的"账号类型"为"企业账号"时，显示该参数，且参数必选。

参数	参数说明
标签	可选参数。 域名的标识，包括键和值，每个域名可以创建10个标签。
描述	可选参数。 域名的描述信息。长度不超过255个字符。

----结束

添加 A 类型记录集

步骤1 [创建内网域名](#)成功后，在“内网域名”列表，单击域名名称“dwr-lms.{regionid}.myhuaweicloud.com”。

步骤2 单击“添加记录集”，进入“添加记录集”页面。

步骤3 在“添加记录集”页面中，设置记录集参数。

更多参数详情请参见[配置内网域名解析](#)。

图 4-4 添加 A 类型记录集

添加记录集 [帮助文档](#)

主机记录

例如www

dwr-lms.cn.myh...

主机记录指域名前缀，例如 example.com 常用的解析如下：
网站解析：主机记录写www，解析的域名是www.example.com
网站解析：主机记录为空，解析的域名是example.com
子域名：主机记录写cdn，解析的域名是cdn.example.com
邮箱解析：主机记录写mail，解析的域名是 mail.example.com
泛解析：主机记录写*，解析的域名是 *.example.com，匹配example.com的所有子域名

* 记录类型

A – 将域名指向IPv4地址

* TTL (秒)

300

* 记录值

192.168.1.1

A记录：填写IPv4地址，最多可以输入50个不重复地址，每行一个。
例如：
192.168.10.10
172.16.100.100

标签

如果您需要使用同一标签标识多种云资源，即所有服务均可在标签输入框下拉选择同一标签，建议在TMS中创建预定义标签。[查看预定义标签](#)
在下方键/值输入框输入内容后单击'添加'，即可将标签加入此处

请输入标签键

请输入标签值

添加

您还可以添加20个标签。

描述

0/255

取消

确定

表 4-3 A 类型记录集参数说明

参数	参数说明
主机记录	解析域名的前缀。 不填。
记录类型	记录集的类型，此处为A类型。 选择“A – 将域名指向IPv4地址”。

文档版本 01 (2025-09-29)

版权所有 © 华为云计算技术有限公司

15

参数	参数说明
TTL(秒)	解析记录在本地DNS服务器的缓存时间，以秒为单位。 默认值为“300秒”。取值范围：1~2147483647。
记录值	域名对应的IPv4地址。 最多可以输入50个不重复地址，多个地址之间以换行符分隔。 说明 这里填写的IP地址是来自于"vpc终端节点基本信息"中的"IPv4地址"中的ip地址
权重	可选参数。 返回解析记录的权重比例。默认值为1，取值范围：0~1000。
标签	可选参数。 记录集的标识，包括键和值，每个记录集可以创建10个标签。
描述	可选参数。 对域名的描述。长度不超过255个字符。

步骤4 单击“确定”，返回“解析记录”页面。添加完成后，您可以在域名对应的记录集列表中查看已添加的记录集。当记录集的状态显示为“正常”时，表示记录集添加成功。

步骤5 请按照以上步骤再次添加主机记录为“*”的A类型记录集。

----结束

 说明

上面步骤都完成后，就可以使用[LMS SDK](#)和[LMS API](#)进行操作了。其中域名就使用上面创建的内网域名。

5查看监控指标

5.1 支持的监控指标

功能说明

本节定义了LMS上报云监控服务的监控指标的命名空间，监控指标列表和维度定义，用户可以通过云监控服务提供管理控制台或API接口来检索LMS产生的监控指标和告警信息。

命名空间

SYS.LMS

监控指标

指标ID	指标名称	含义	取值范围	测量对象&维度	监控周期（原始指标）
collection_count	集合总数	该指标用于统计知识仓中的集合总数。 单位：count	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟

指标ID	指标名称	含义	取值范围	测量对象&维度	监控周期（原始指标）
vector_count	向量总数	该指标用于统计知识仓实例中向量的总数。 单位：count	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟
total_capacity	总存储容量	该指标用于统计知识仓实例所在存储总容量。 单位：Byte	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟
insert_request_qps	Insert请求QPS	该指标用于统计每秒插入向量请求次数。 单位：count/s	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟
upsert_request_qps	Upsert请求QPS	该指标用于统计每秒Upsert向量请求次数。 单位：count/s	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟
delete_request_qps	Delete请求QPS	该指标用于统计每秒删除向量请求次数。 单位：count/s	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟

指标ID	指标名称	含义	取值范围	测量对象&维度	监控周期（原始指标）
search_request_qps	Search请求QPS	该指标用于统计每秒搜索向量请求次数。 单位：count/s	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟
query_request_qps	Query请求QPS	该指标用于统计每秒查询标量请求次数。 单位：count/s	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟
hybrid_search_request_qps	Hybrid_Search请求QPS	该指标用于统计每秒混合搜索请求次数。 单位：count/s	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟
insert_request_latency_avg	Insert请求平均时延	该指标用于统计周期内插入向量请求平均时延。 单位：ms	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟
upsert_request_latency_avg	Upsert请求平均时延	该指标用于统计周期内Upsert向量请求平均时延。 单位：ms	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟

指标ID	指标名称	含义	取值范围	测量对象&维度	监控周期（原始指标）
delete_request_latency_avg	Delete请求平均时延	该指标用于统计周期内删除向量请求平均时延。 单位：ms	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟
search_request_latency_avg	Search请求平均时延	该指标用于统计周期内搜索向量请求平均时延。 单位：ms	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟
query_request_latency_avg	Query请求平均时延	该指标用于统计周期内查询标量请求平均时延。 单位：ms	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟
hybrid_search_request_latency_avg	Hybrid_Search请求平均时延	该指标用于统计周期内混合搜索请求平均时延。 单位：ms	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟
insert_request_latency_p99	Insert请求p99时延	该指标用于统计插入向量请求的p99时延。 单位：ms	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟

指标ID	指标名称	含义	取值范围	测量对象&维度	监控周期（原始指标）
upsert_request_latency_p99	Upsert请求p99时延	该指标用于统计Upsert向量请求的p99时延。 单位：ms	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟
delete_request_latency_p99	Delete请求p99时延	该指标用于统计删除向量请求的p99时延。 单位：ms	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟
search_request_latency_p99	Search请求p99时延	该指标用于统计搜索向量请求的p99时延。 单位：ms	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟
query_request_latency_p99	Query请求p99时延	该指标用于统计查询标量请求的p99时延。 单位：ms	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟
hybrid_search_request_latency_p99	Hybrid_Search请求p99时延	该指标用于统计混合搜索请求的p99时延。 单位：ms	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟

指标ID	指标名称	含义	取值范围	测量对象&维度	监控周期（原始指标）
request_success_rate	请求成功率	该指标用于衡量客户端请求的成功率。成功请求占总请求数的百分比，计算方式：(请求成功数量/总请求数量)*100%。 单位：%	≥ 0 , ≤ 100	测量对象： 知识仓 测量维度： store_instance_id	1分钟
insert_vector_count	Insert向量数	该指标用于统计插入向量总数。 单位：count	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟
upsert_vector_count	Upsert向量数	该指标用于统计Upsert向量总数。 单位：count	≥ 0	测量对象： 知识仓 测量维度： store_instance_id	1分钟
compute_resource_rate	计算资源利用率	该指标用于统计已使用的计算资源占总计算资源的百分比，计算方式：(已使用的计算资源/总计算资源)*100%。 单位：%	≥ 0 , ≤ 100	测量对象： 知识仓 测量维度： store_instance_id	1分钟
memory_resource_rate	内存容量利用率	该指标用于统计已使用的内存容量占总内存容量的百分比，计算方式：(已使用的内存容量/总内存容量)*100%。 单位：%	≥ 0 , ≤ 100	测量对象： 知识仓 测量维度： store_instance_id	1分钟