

DataArtsFabric

API 参考

文档版本 01
发布日期 2025-09-15



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 使用前必读..... 1
- 2 API 概览..... 3
- 3 如何调用 API.....4
 - 3.1 构造请求.....4
 - 3.2 认证鉴权.....7
 - 3.3 返回结果.....8
- 4 API..... 10
 - 4.1 服务健康检查.....10
 - 4.1.1 健康检查.....10
 - 4.2 用户协议管理.....15
 - 4.2.1 查询系统协议.....15
 - 4.2.2 注册租户协议.....21
 - 4.2.3 查询用户是否注册协议.....26
 - 4.2.4 删除用户注册协议.....31
 - 4.3 委托管理.....36
 - 4.3.1 创建服务委托.....36
 - 4.3.2 查询服务委托.....42
 - 4.3.3 删除服务委托.....49
 - 4.4 工作空间管理.....55
 - 4.4.1 查询工作空间列表.....55
 - 4.4.2 更新工作空间.....63
 - 4.4.3 创建工作空间.....71
 - 4.4.4 删除工作空间.....79
 - 4.5 服务规格.....84
 - 4.5.1 查询资源规格列表.....84
 - 4.6 端点管理.....92
 - 4.6.1 查询端点列表.....92
 - 4.6.2 创建端点.....107
 - 4.6.3 查询端点详情.....127
 - 4.6.4 修改端点.....134
 - 4.6.5 删除端点.....153
 - 4.6.6 订阅端点.....158

4.7 模型定义.....	163
4.7.1 创建模型.....	163
4.7.2 查询模型列表.....	171
4.7.3 列举基模型.....	182
4.7.4 删除模型的指定版本.....	188
4.7.5 更新模型信息.....	193
4.7.6 查看模型版本列表.....	202
4.7.7 删除未使用的模型定义.....	210
4.8 TMS 管理.....	215
4.8.1 查询项目标签.....	216
4.8.2 查询资源标签.....	221
4.8.3 查询资源实例列表.....	228
4.8.4 查询资源实例数量.....	237
4.8.5 批量添加资源标签.....	244
4.8.6 批量删除资源标签.....	251
4.9 消息管理.....	257
4.9.1 创建消息通知策略.....	257
4.9.2 查询消息通知策略列表.....	264
4.9.3 删除消息通知策略.....	271
4.10 监控管理.....	276
4.10.1 更新 AOM 监控采集配置.....	276
4.11 接入客户端.....	281
4.11.1 创建 vpcep 客户端.....	281
4.11.2 查询 vpcep 客户端列表.....	290
4.11.3 查询 vpcep 客户端详情.....	298
4.11.4 删除 vpcep 客户端.....	305
4.11.5 更新 vpcep 客户端.....	309
4.12 Service 实例管理.....	314
4.12.1 列举已部署的服务实例.....	315
4.12.2 查看部署的服务实例详情.....	327
4.12.3 部署服务.....	333
4.12.4 更新已部署的服务实例.....	341
4.12.5 删除服务实例.....	349
4.13 推理服务接口.....	354
4.13.1 发起调用请求.....	354
4.14 Ray Session 接口.....	362
4.14.1 获取 Session 信息.....	363
4.15 SQL 业务接口.....	368
4.15.1 创建 SQL 的 Session.....	368
4.15.2 获取 SQL Session 列表.....	374
4.15.3 查询 Session 信息.....	382
4.15.4 关闭 Session.....	387

4.15.5 执行 SQL.....	392
4.15.6 获取 Statement 结果、执行状态.....	399
4.15.7 取消 Statement 执行.....	405
4.15.8 获取 SQL 执行记录.....	411
4.15.9 获取指定 Session 下的 SQL 执行记录.....	419
5 应用示例.....	427
5.1 创建一个推理服务.....	427
6 权限和授权项.....	430
6.1 权限及授权项说明.....	430
6.2 策略授权参考.....	431
7 附录.....	438
7.1 状态码.....	438
7.2 错误码.....	440
7.3 获取项目 ID.....	449
7.4 获取账号 ID.....	450
7.5 获取工作空间 ID.....	450
7.6 获取端点 ID.....	451
7.7 获取模型 ID.....	452
7.8 获取推理服务实例 ID.....	452
7.9 获取企业项目 ID.....	453

1 使用前必读

欢迎使用Fabric服务。您可以通过控制台或API实现大模型推理、Ray作业等服务。

您可以使用本文档提供的API对数智融合计算服务进行相关操作，例如创建删除推理、作业端点等。支持的全部操作请参见[API概览](#)。

终端节点

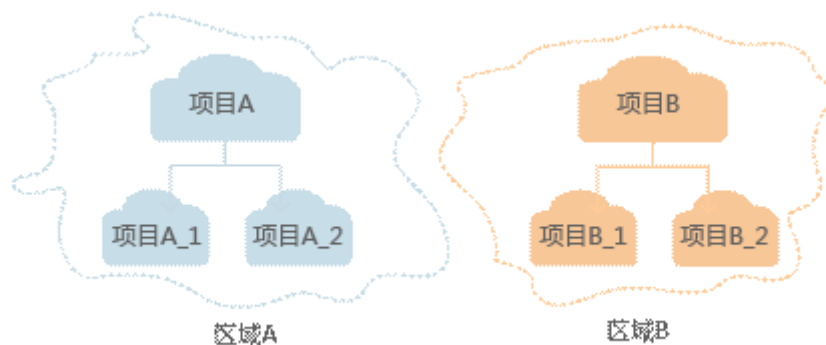
终端节点即调用API的[请求地址](#)，不同服务不同区域的终端节点不同，您可以从[地区和终端节点](#)中查询服务的终端节点。请您根据业务需要选择对应区域的终端节点。

基本概念

- 账号
用户注册华为云时的账号，账号对其所拥有的资源及云服务具有完全的访问权限，可以重置用户密码、分配用户权限等。由于账号是付费主体，为了确保账号安全，建议您不要直接使用账号进行日常管理工作，而是创建用户并用于日常管理工作。
- 用户
由账号在IAM中创建的用户，是云服务的使用人员，具有身份凭证（密码和访问密钥）。
在[我的凭证](#)下，您可以查看账号ID和用户ID。通常在调用API的鉴权过程中，您需要用到账号、用户和密码等信息。
- 区域（Region）
从地理位置和网络时延维度划分，同一个Region内共享弹性计算、块存储、对象存储、VPC网络、弹性公网IP、镜像等公共服务。Region分为通用Region和专属Region，通用Region指面向公共租户提供通用云服务的Region；专属Region指只承载同一类业务或只面向特定租户提供业务服务的专用Region。详细信息，请参见[区域和可用区](#)。
- 可用区（AZ，Availability Zone）
一个AZ是一个或多个物理数据中心的集合，有独立的风火水电，AZ内逻辑上再将计算、网络、存储等资源划分成多个集群。一个Region中的多个AZ间通过高速光纤相连，以满足用户跨AZ构建高可用性系统的需求。
- 项目
华为云的区域默认对应一个项目，这个项目由系统预置，用来隔离物理区域间的资源（计算资源、存储资源和网络资源），以默认项目为单位进行授权，用户可

以访问您账号中该区域的所有资源。如果您希望进行更加精细的权限控制，可以在区域默认的项目中创建子项目，并在子项目中购买资源，然后以子项目为单位进行授权，使得用户仅能访问特定子项目中资源，使得资源的权限控制更加精确。您可以在[我的凭证](#)下，查看项目ID。

图 1-1 项目隔离模型



- 企业项目

企业项目是项目的升级版，针对企业不同项目间资源的分组和管理，是逻辑隔离。企业项目中可以包含多个区域的资源，且项目中的资源可以迁入迁出。

关于企业项目特性的详细信息，请参见[企业管理服务](#)。

2 API 概览

通过使用Fabric服务所提供的接口，您可以完整的使用Fabric服务的基本功能。Fabric服务提供的具体API如表2-1所示。

表 2-1 Fabric 服务 API 概览

类型	子类型	说明
Fabric-admin	服务健康检查	服务健康检查。
	用户协议管理	用户协议管理。
	委托管理	委托管理。
	Workspace管理	Workspace管理。
	服务规格	服务规格。
	端点管理	端点管理。
	模型定义	模型定义。
	TMS管理	TMS管理。
	消息管理	消息管理。
	监控管理	监控管理。
	配置中心	配置中心。
Fabric-ep	列举已部署的服务实例	列举已部署的服务实例。
	查看部署的服务实例详情	查看部署的服务实例详情。
	部署服务	部署服务。
	更新已部署的服务实例	更新已部署的服务实例。
	删除服务实例	删除服务实例。
	发起调用请求	发起调用请求。

3 如何调用 API

3.1 构造请求

本节介绍REST API请求的组成，并以调用IAM服务的[管理员创建IAM用户](#)接口说明如何调用API。

请求 URI

请求URI由如下部分组成。

{URI-scheme} :// {Endpoint} / {resource-path} ? {query-string}

尽管请求URI包含在请求消息头中，但大多数语言或框架都要求您从请求消息中单独传递它，所以在此单独强调。

- **URI-scheme**：表示用于传输请求的协议，当前所有API均采用HTTPS协议。
- **Hostname**：指定承载REST服务端点的服务器域名或IP，不同服务不同区域的Hostname不同，您可以从[地区和终端节点](#)中获取。例如IAM服务在“中国-香港”区域的Hostname为“iam.ap-southeast-1.myhuaweicloud.com”。
- **resource-path**：资源路径，也即API访问路径。从具体API的URI模块获取，例如“管理员创建IAM用户”接口的resource-path为“/v3.0/OS-USER/users”。
- **query-string**：查询参数，是可选部分，并不是每个API都有查询参数。查询参数前面需要带一个“？”，形式为“参数名=参数取值”，例如“limit=10”，表示查询不超过10条数据。

例如您需要创建IAM用户，由于IAM为全局服务，则使用任一区域的Endpoint（例如“中国-香港”区域的Hostname：“iam.ap-southeast-1.myhuaweicloud.com”），并在[管理员创建IAM用户](#)的URI部分找到resource-path（/v3.0/OS-USER/users），拼接起来如下所示。

```
https://iam.ap-southeast-1.myhuaweicloud.com/v3.0/OS-USER/users
```

说明

为查看方便，在每个具体API的URI部分，只给出resource-path部分，并将请求方法写在一起。这是因为URI-scheme都是HTTPS，而Hostname在同一个区域也相同，所以简洁起见将这两部分省略。

请求方法

HTTP请求方法（也称为操作或动词），它告诉服务正在请求什么类型的操作。

表 3-1 请求方法说明

请求方法	说明
GET	请求服务器返回指定资源。
PUT	请求服务器更新指定资源。
POST	请求服务器新增资源或执行特殊操作。
DELETE	请求服务器删除指定资源，例如删除对象等。
HEAD	请求服务器资源头部。
PATCH	请求服务器更新资源的部分内容。当资源不存在的时候，PATCH可能会去创建一个新的资源。

在[管理员创建IAM用户](#)的URI部分，您可以看到其请求方法为“POST”，则其请求为：

```
POST https://iam.ap-southeast-1.myhuaweicloud.com/v3.0/OS-USER/users
```

请求消息头

附加请求头字段，如指定的URI和HTTP方法所要求的字段。例如定义消息体类型的请求头“Content-Type”，请求鉴权信息等。

如下公共消息头需要添加到请求中。

表 3-2 公共消息头说明

字段	是否必选	说明
Content-Type	是	消息体的类型（格式）。默认值为“application/json”，有其他取值时会在具体接口中专门说明。
Authorization	否	签名认证信息。当使用AK/SK方式认证时，使用SDK对请求进行签名的过程中会自动填充该字段。AK/SK认证的详细说明请参考 AK/SK认证 。
X-Sdk-Date	否	请求发送的时间。当使用AK/SK方式认证时，使用SDK对请求进行签名的过程中会自动填充该字段。AK/SK认证的详细说明请参考 AK/SK认证 。
X-Auth-Token	否	用户Token。当使用Token方式认证时，必须填充该字段。用户Token也就是调用 获取用户Token 接口的响应值，该接口是唯一不需要认证的接口。

字段	是否必选	说明
X-Project-ID	否	子项目ID。在多项目场景中使用。如果云服务资源创建在子项目中，AK/SK认证方式下，操作该资源的接口调用需要在请求消息头中携带X-Project-ID。
X-Domain-ID	否	账号ID。AK/SK认证方式下，全局服务的接口调用时，需在请求消息头中携带X-Domain-ID。

对于[管理员创建IAM用户](#)接口，使用AK/SK方式认证时，添加消息头后的请求如下所示。

```
POST https://iam.ap-southeast-1.myhuaweicloud.com/v3.0/OS-USER/users
Content-Type: application/json
X-Sdk-Date: 20240416T095341Z
Authorization: SDK-HMAC-SHA256 Access=*****, SignedHeaders=content-type;host;x-sdk-date,
Signature=*****
```

请求消息体

请求消息体通常以结构化格式发出，与请求消息头中Content-type对应，传递除请求消息头之外的内容。如果请求消息体中参数支持中文，则中文字符必须为UTF-8编码，并在Content-type中声明字符编码方式，例如：Content-Type: application/json; charset=utf-8。

每个接口的请求消息体内容不同，也并不是每个接口都需要有请求消息体（或者说消息体为空），GET、DELETE操作类型的接口就不需要消息体，消息体具体内容需要根据具体接口而定。

对于[管理员创建IAM用户](#)接口，您可以从接口的请求部分看到所需的请求参数及参数说明，将消息体加入后的请求如下所示，其中加粗的字段需要根据实际值填写。

表 3-3 字段说明

字段	说明
accountid	IAM用户所属的账号ID。
username	要创建的IAM用户名。
email	IAM用户的邮箱。
*****	IAM用户的登录密码。

```
POST https://iam.ap-southeast-1.myhuaweicloud.com/v3.0/OS-USER/users
Content-Type: application/json
X-Sdk-Date: 20240416T095341Z
Authorization: SDK-HMAC-SHA256 Access=*****, SignedHeaders=content-type;host;x-sdk-date,
Signature=*****

{
  "user": {
    "domain_id": "accountid",
    "name": "username",
```

```
"password": "*****",
"email": "email",
"description": "IAM User Description"
}
```

到此为止，一个API请求所需要的内容已经准备完成。您可以使用curl、Postman或直接编写代码等方式发送请求调用API。

3.2 认证鉴权

调用接口有如下两种认证方式，您可以选择其中一种进行认证鉴权。

- AK/SK认证：通过AK（Access Key ID）/SK（Secret Access Key）加密调用请求。
- Token认证：通过Token认证调用请求。

AK/SK 认证

说明

- AK/SK签名认证方式仅支持消息体大小12MB以内，12MB以上的请求请使用Token认证。
- AK/SK既可以使用永久访问密钥中的AK/SK，也可以使用临时访问密钥中的AK/SK，但使用临时访问密钥的AK/SK时需要额外携带“X-Security-Token”字段，字段值为临时访问密钥的security_token。

AK/SK认证就是使用AK/SK对请求进行签名，在请求时将签名信息添加到消息头，从而通过身份认证。

- AK（Access Key ID）：访问密钥ID。与私有访问密钥关联的唯一标识符；访问密钥ID和私有访问密钥一起使用，对请求进行加密签名。
- SK（Secret Access Key）：与访问密钥ID结合使用的密钥，对请求进行加密签名，可标识发送方，并防止请求被修改。

使用AK/SK认证时，您可以基于签名算法使用AK/SK对请求进行签名，也可以使用专门的签名SDK对请求进行签名。详细的签名方法和SDK使用方法请参见[API签名指南](#)。

须知

签名SDK只提供签名功能，与服务提供的SDK不同，使用时请注意。

Token 认证

说明

- Token的有效期为24小时，需要使用一个Token鉴权时，可以先缓存起来，避免频繁调用。
- 使用Token前请确保Token离过期有足够的时间，防止调用API的过程中Token过期导致调用API失败。

Token在计算机系统中代表令牌（临时）的意思，拥有Token就代表拥有某种权限。Token认证就是在调用API的时候将Token加到请求消息头，从而通过身份认证，获得操作API的权限。

Token可通过调用[获取用户Token](#)接口获取，调用本服务API需要project级别的Token，即调用[获取用户Token](#)接口时，请求body中auth.scope的取值需要选择project，如下所示。

```
{
  "auth": {
    "identity": {
      "methods": [
        "password"
      ],
      "password": {
        "user": {
          "name": "username",
          "password": "*****#",
          "domain": {
            "name": "domainname"
          }
        }
      }
    },
    "scope": {
      "project": {
        "name": "xxxxxxx"
      }
    }
  }
}
```

获取Token后，再调用其他接口时，您需要在请求消息头中添加“X-Auth-Token”，其值即为获取到的Token。例如Token值为“ABCDEFGH....”，则调用接口时将“X-Auth-Token: ABCDEFG....”加到请求消息头即可，如下所示。

```
POST https://iam.ap-southeast-1.myhuaweicloud.com/v3.0/OS-USER/users
Content-Type: application/json
X-Auth-Token: ABCDEFG....
```

3.3 返回结果

请求发送以后，您会收到响应，包含状态码、响应消息头和消息体。

状态码

状态码是一组从1xx到5xx的数字代码，状态码表示了请求响应的状态，完整的状态码列表请参见[状态码](#)。

对于[管理员创建IAM用户](#)接口，如果调用后返回状态码为“201”，则表示请求成功。

响应消息头

对应请求消息头，响应同样也有消息头，如“Content-type”。

对于[管理员创建IAM用户](#)接口，返回如[图3-1](#)所示的消息头。

图 3-1 管理员创建 IAM 用户响应消息头

```
"X-Frame-Options": "SAMEORIGIN",
"X-IAM-ETag-id": "2562365939-d8f6f12921974cb097338ac11fceac8a",
"Transfer-Encoding": "chunked",
"Strict-Transport-Security": "max-age=31536000; includeSubdomains;",
"Server": "api-gateway",
"X-Request-Id": "af2953f2bcc67a42325a69a19e6c32a2",
"X-Content-Type-Options": "nosniff",
"Connection": "keep-alive",
"X-Download-Options": "noopen",
"X-XSS-Protection": "1; mode=block;",
"X-IAM-Trace-Id": "token_ _null_af2953f2bcc67a42325a69a19e6c32a2",
"Date": "Tue, 21 May 2024 09:03:40 GMT",
"Content-Type": "application/json; charset=utf8"
```

响应消息体

响应消息体通常以结构化格式返回，与响应消息头中Content-type对应，传递除响应消息头之外的内容。

对于[管理员创建IAM用户](#)接口，返回如下消息体。为篇幅起见，这里只展示部分内容。

```
{
  "user": {
    "id": "c131886aec...",
    "name": "IAMUser",
    "description": "IAM User Description",
    "areacode": "",
    "phone": "",
    "email": "****@***.com",
    "status": null,
    "enabled": true,
    "pwd_status": false,
    "access_mode": "default",
    "is_domain_owner": false,
    "xuser_id": "",
    "xuser_type": "",
    "password_expires_at": null,
    "create_time": "2024-05-21T09:03:41.000000",
    "domain_id": "d78cbac1.....",
    "xdomain_id": "30086000.....",
    "xdomain_type": "",
    "default_project_id": null
  }
}
```

当接口调用出错时，会返回错误码及错误信息说明，错误响应的Body体格式如下所示。

```
{
  "error_msg": "Request body is invalid.",
  "error_code": "IAM.0011"
}
```

其中，error_code表示错误码，error_msg表示错误描述信息。

4 API

4.1 服务健康检查

4.1.1 健康检查

功能介绍

查看系统健康状态，用户可通过该接口查看系统健康状态。此接口为同步接口，无配套使用接口和特殊场景。

URI

GET /healthcheck

请求参数

无

响应参数

状态码：200

OK

状态码：400

表 4-1 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401**表 4-2 响应 Body 参数**

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404**表 4-3 响应 Body 参数**

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408**表 4-4 响应 Body 参数**

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500**表 4-5 响应 Body 参数**

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：503

表 4-6 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

GET https://{endpoint}/healthcheck

响应示例

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码：503

ServiceUnavailable

```
{
  "error_code" : "common.00000503",
  "error_msg" : "service unavailable"
}
```

状态码

状态码	描述
200	OK

状态码	描述
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError
503	ServiceUnavailable

错误码

请参见[错误码](#)。

4.2 用户协议管理

4.2.1 查询系统协议

功能介绍

查询租户协议内容，用户可从中了解协议名称、版本等信息。该接口为同步接口，无配套使用接口和特殊场景。

URI

GET /v1/agreement-rule

请求参数

表 4-7 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

响应参数

状态码：200

表 4-8 响应 Body 参数

参数	参数类型	描述
agreement_rules	Array of AgreementRule objects	参数解释：系统协议列表。 取值范围：列表长度为[0-1000]。

表 4-9 AgreementRule

参数	参数类型	描述
agreement_name	String	参数解释：协议名称。 约束限制：不涉及。 取值范围：长度不超过65534个字符。 默认取值：不涉及。
agreement_display_name	String	参数解释：协议展示名称。 约束限制：不涉及。 取值范围：长度不超过65534个字符。 默认取值：不涉及。
agreement_version	String	参数解释：协议版本。 约束限制：不涉及。 取值范围：长度不超过65534个字符。 默认取值：不涉及。
agreement_url	String	参数解释：协议链接。 约束限制：不涉及。 取值范围：长度不超过65534个字符。 默认取值：不涉及。

状态码：400

表 4-10 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401**表 4-11 响应 Body 参数**

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404**表 4-12 响应 Body 参数**

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-13 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-14 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

/v1/agreement-rule

响应示例

状态码：200

系统协议响应体。

```
{
  "agreement_rules": [ {
    "agreement_name": "agreement AAA",
    "agreement_display_name": "AAA",
    "agreement_version": "v1",
    "agreement_url": "https://AAA"
  }, {
    "agreement_name": "agreement BBB",
    "agreement_display_name": "BBB",
    "agreement_version": "v1",
    "agreement_url": "https://BBB"
  } ]
}
```

状态码：400

BadRequest

```
{
  "error_code": "common.01000001",
  "error_msg": "failed to read http request, please check your input, code: 400, reason: Type mismatch., cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code": "APIG.1002",
  "error_msg": "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error": {
```

```
"code" : "403",
"message" : "X-Auth-Token is invalid in the request",
"title" : "Forbidden"
},
"error_code" : 403,
"error_msg" : "X-Auth-Token is invalid in the request",
"title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	系统协议响应体。
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.2.2 注册租户协议

功能介绍

注册租户协议，用户通过该接口注册租户协议。此接口为同步接口，无配套使用接口和特殊场景。

URI

POST /v1/agreement

请求参数

表 4-15 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

表 4-16 请求 Body 参数

参数	是否必选	参数类型	描述
agreements	否	Array of TenantAgreement objects	参数解释： 租户协议列表。 约束限制： [0,1000]。

表 4-17 TenantAgreement

参数	是否必选	参数类型	描述
agreement_name	否	String	参数解释： 协议名称。 约束限制： 不涉及。 取值范围： 长度为[1,64]的字母、数字、下划(_)和中划线(-)的组合。 默认取值： 不涉及。

参数	是否必选	参数类型	描述
agreement_version	否	String	参数解释： 协议版本号。 约束限制： 不涉及。 取值范围： 长度为[1,32]的字母、数字的组合。 默认取值： 不涉及。

响应参数

状态码：200

表 4-18 响应 Body 参数

参数	参数类型	描述
-	AnyType	OK

状态码：400

表 4-19 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释： 错误码。 约束限制： 不涉及。 取值范围： 长度为[8,36]个字符。 默认取值： 不涉及。
error_msg	String	参数解释： 错误描述。 约束限制： 不涉及。 取值范围： 长度为[2,512]个字符。 默认取值： 不涉及。
solution_msg	String	参数解释： 解决方案描述。 约束限制： 不涉及。 取值范围： 长度为[0,4096]个字符。 默认取值： 不涉及。

状态码：401

表 4-20 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-21 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-22 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-23 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

用户通过该接口注册租户协议，协议名称为test_name，协议版本为v1。

```
/v1/agreement
{
```

```
"agreements": [ {  
  "agreement_name": "test_name",  
  "agreement_version": "v1"  
}]  
}
```

响应示例

状态码：400

BadRequest

```
{  
  "error_code": "common.01000001",  
  "error_msg": "failed to read http request, please check your input, code: 400, reason: Type mismatch.,  
cause: TypeMismatchException"  
}
```

状态码：401

Unauthorized

```
{  
  "error_code": "APIG.1002",  
  "error_msg": "Incorrect token or token resolution failed"  
}
```

状态码：403

Forbidden

```
{  
  "error": {  
    "code": "403",  
    "message": "X-Auth-Token is invalid in the request",  
    "title": "Forbidden"  
  },  
  "error_code": 403,  
  "error_msg": "X-Auth-Token is invalid in the request",  
  "title": "Forbidden"  
}
```

状态码：404

NotFound

```
{  
  "error_code": "common.01000001",  
  "error_msg": "response status exception, code: 404"  
}
```

状态码：408

Request Time-out

```
{  
  "error_code": "common.00000408",  
  "error_msg": "timeout exception occurred"  
}
```

状态码：500

InternalServerError

```
{  
  "error_code": "common.00000500",  
  "error_msg": "internal error"  
}
```

状态码

状态码	描述
200	OK
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.2.3 查询用户是否注册协议

功能介绍

查询用户是否注册协议。此接口为同步接口，无配套使用接口和特殊场景。

URI

GET /v1/agreement

请求参数

表 4-24 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

响应参数

状态码：200

表 4-25 响应 Body 参数

参数	参数类型	描述
agreements	Array of TenantAgreement objects	参数解释：租户协议列表。 约束限制：不涉及。 取值范围：[0,1000]。 默认取值：不涉及。

表 4-26 TenantAgreement

参数	参数类型	描述
agreement_name	String	参数解释：协议名称。 约束限制：不涉及。 取值范围：长度为[1,64]的字母、数字、下划(_)和中划线(-)的组合。 默认取值：不涉及。
agreement_version	String	参数解释：协议版本号。 约束限制：不涉及。 取值范围：长度为[1,32]的字母、数字的组合。 默认取值：不涉及。

状态码：400

表 4-27 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401**表 4-28 响应 Body 参数**

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404**表 4-29 响应 Body 参数**

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-30 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-31 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

GET https://{endpoint}/v1/agreement

响应示例

状态码：200

租户注册协议响应体。

```
{
  "agreements": [ {
    "agreement_name": "test_name",
    "agreement_version": "v1"
  } ]
}
```

状态码：400

BadRequest

```
{
  "error_code": "common.01000001",
  "error_msg": "failed to read http request, please check your input, code: 400, reason: Type mismatch., cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code": "APIG.1002",
  "error_msg": "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error": {
    "code": "403",
    "message": "X-Auth-Token is invalid in the request",
    "title": "Forbidden"
  },
  "error_code": 403,
  "error_msg": "X-Auth-Token is invalid in the request",
  "title": "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code": "common.01000001",
  "error_msg": "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code": "common.00000408",
  "error_msg": "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	租户注册协议响应体。
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.2.4 删除用户注册协议

功能介绍

删除用户注册协议。此接口为同步接口，无配套使用接口和特殊场景。

URI

DELETE /v1/agreement

请求参数

表 4-32 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

响应参数

状态码：200

表 4-33 响应 Body 参数

参数	参数类型	描述
-	AnyType	OK

状态码：400

表 4-34 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-35 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404**表 4-36** 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408**表 4-37** 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-38 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

GET https://{endpoint}/v1/agreement

响应示例

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	OK
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out

状态码	描述
500	InternalServerError

错误码

请参见[错误码](#)。

4.3 委托管理

4.3.1 创建服务委托

功能介绍

授权委托给Fabric服务。用户调用该接口会创建一个委托，该委托会授权给Fabric服务。用户可以通过传入支持的委托权限策略来可选择的授权。此接口为同步接口，无配套使用接口和特殊场景。

URI

POST /v1/agency

请求参数

表 4-39 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户Token。用于调用接口的认证。获取方法，请参见认证鉴权。 约束限制：不涉及。 取值范围：长度不超过65534个字符。 默认取值：不涉及。

表 4-40 请求 Body 参数

参数	是否必选	参数类型	描述
policy_types	否	Array of strings	参数解释：权限类型列表，使用 DataArtsFabric 的功能，可能需要授权一些权限策略，可以在 policy_types 中加入策略类型来授权权限。 约束限制：不涉及。 取值范围： • FABRIC_COMMON_POLICY：基础通用服务权限策略，是委托必需的权限策略； • FABRIC_SELF_POLICY：访问自身服务所需要的权限策略； • FABRIC_SMN_POLICY：消息通知功能相关权限策略，用来将系统通知消息转发到 SMN； • FABRIC_LAKEFORMATION_POLICY：LakeFormation 功能相关权限策略； • FABRIC_AOM_POLICY：AOM 功能相关权限策略。 • FABRIC_SWR_POLICY：镜像包注册功能相关权限策略。 • FABRIC_LTS_POLICY：日志功能相关权限策略。 • FABRIC_VPCEP_POLICY：网络打通功能相关权限策略。 • FABRIC_DNS_POLICY：DNS 相关权限策略。 • FABRIC_OBS_POLICY：OBS 相关权限策略。 默认取值：不涉及。

响应参数

状态码：200

表 4-41 响应 Body 参数

参数	参数类型	描述
-	AnyType	OK

状态码：400

表 4-42 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-43 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-44 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-45 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-46 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

授权FABRIC_LAKEFORMATION_POLICY策略。

```
POST https://{endpoint}/v1/agency
{
  "policy_types" : [ "FABRIC_LAKEFORMATION_POLICY" ]
}
```

响应示例

状态码：200

OK

```
{
  "policy_types" : [ "FABRIC_LAKEFORMATION_POLICY" ]
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error": {
    "code": "403",
    "message": "X-Auth-Token is invalid in the request",
    "title": "Forbidden"
  },
  "error_code": 403,
  "error_msg": "X-Auth-Token is invalid in the request",
  "title": "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code": "common.01000001",
  "error_msg": "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code": "common.00000408",
  "error_msg": "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code": "common.00000500",
  "error_msg": "internal error"
}
```

状态码

状态码	描述
200	OK
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.3.2 查询服务委托

功能介绍

查询授权给DataArtsFabric服务的委托。用户调用该接口查询已授权给DataArtsFabric的委托，以及委托中具体的权限内容。此接口为同步接口，无配套使用接口和特殊场景。

URI

GET /v1/agency

表 4-47 Query 参数

参数	是否必选	参数类型	描述
policy_type	否	String	参数解释：策略类型。 约束限制：支持模糊匹配。 取值范围： • FABRIC_COMMON_POLICY：基础通用服务相关权限策略，是委托必需的权限策略； • FABRIC_SELF_POLICY：访问自身服务所需要的权限策略； • FABRIC_SMN_POLICY：消息通知功能相关权限策略，用来将系统通知消息转发到SMN； • FABRIC_LAKEFORMATION_POLICY：LakeFormation功能相关权限策略； • FABRIC_AOM_POLICY：AOM功能相关权限策略。 • FABRIC_SWR_POLICY：镜像包注册功能相关权限策略。 • FABRIC_LTS_POLICY：日志功能相关权限策略。 • FABRIC_VPCEP_POLICY：网络打通功能相关权限策略。 • FABRIC_DNS_POLICY：DNS相关权限策略。 • FABRIC_OBS_POLICY：OBS相关权限策略。 默认取值：不涉及。

请求参数

表 4-48 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户Token。用于调用接口的认证。获取方法，请参见认证鉴权。 约束限制：不涉及。 取值范围：长度不超过65534个字符。 默认取值：不涉及。

响应参数

状态码：200

表 4-49 响应 Body 参数

参数	参数类型	描述
agency_name	String	参数解释：委托名称。 约束限制：不涉及。 取值范围：长度[1,128]。 默认取值：不涉及。
authorized	Boolean	参数解释：授权与否。如果用户的服务委托权限包含系统所需要的委托权限返回true，否则返回false。 约束限制：不涉及。 取值范围：true，false。 默认取值：不涉及。
agency_policies	Array of AgencyPolicy objects	参数解释：委托策略列表。 约束限制：[1,100]。

表 4-50 AgencyPolicy

参数	参数类型	描述
policy_type	String	参数解释：权限类型列表；使用Fabric的功能，可能需要授权一些权限策略，可以在policy_types中加入策略类型来授权权限。 约束限制：不涉及。 取值范围： • FABRIC_COMMON_POLICY：基础通用服务相关权限策略，是委托必需的权限策略； • FABRIC_SELF_POLICY：访问自身服务所需要的权限策略； • FABRIC_SMN_POLICY：消息通知功能相关权限策略，用来将系统通知消息转发到SMN； • FABRIC_LAKEFORMATION_POLICY：LakeFormation功能相关权限策略； • FABRIC_AOM_POLICY：AOM功能相关权限策略。 • FABRIC_SWR_POLICY：镜像包注册功能相关权限策略。 • FABRIC_LTS_POLICY：日志功能相关权限策略。 • FABRIC_VPCEP_POLICY：网络打通功能相关权限策略。 • FABRIC_DNS_POLICY：DNS相关权限策略。 默认取值：不涉及。
is_required	Boolean	参数解释：是否是必需的。 约束限制：不涉及。 取值范围：true, false。 默认取值：不涉及。
is_enable	Boolean	参数解释：是否已开启。 约束限制：不涉及。 取值范围：true, false。 默认取值：不涉及。
actions	Array of strings	参数解释：权限列表。 约束限制：[0,1000]。

参数	参数类型	描述
description	String	参数解释：策略描述。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

状态码：400

表 4-51 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-52 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-53 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-54 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-55 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

GET https://{endpoint}/v1/agency

响应示例

状态码：200

查询委托响应体。

```
{
  "agency_name": "admin_trust",
  "authorized": true,
  "agency_policies": [ {
    "policy_type": "FABRIC_COMMON_POLICY",
    "is_required": false,
    "is_enable": true,
    "actions": [ "lakeformation:instance:show" ],
    "description": "使用DataArtsFabric过程中需要获取一些必要的权限。"
  } ]
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	查询委托响应体。
400	BadRequest

状态码	描述
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.3.3 删除服务委托

功能介绍

取消授权给DataArtsFabric服务的委托。委托中包含由了DataArtsFabric服务定义的必要权限策略和非必要权限策略。非必要权限策略可以通过指定权限策略名称删除，不会影响其他权限策略。必要权限策略不能单独删除，只能通过删除整个委托的方式来删除。部分权限策略删除前会校验是否满足删除条件，不满足删除条件无法删除，以此避免误删权限策略影响业务。此接口为同步接口，无配套使用接口和特殊场景。

URI

DELETE /v1/agency

表 4-56 Query 参数

参数	是否必选	参数类型	描述
policy_type	否	String	参数解释：权限策略类型。 约束限制：支持模糊匹配。 取值范围： ● FABRIC_COMMON_POLICY：基础通用服务相关的权限策略； ● FABRIC_SMN_POLICY：消息通知功能相关的权限策略； ● FABRIC_LAKEFORMATION_POLICY：LakeFormation功能相关的权限策略； ● FABRIC_AOM_POLICY：AOM功能相关的权限策略； ● FABRIC_SELF_POLICY：访问自身服务所需要的权限策略。 ● FABRIC_SWR_POLICY：镜像包注册功能所需要的权限策略。 ● FABRIC_LTS_POLICY：日志功能相关权限策略。 ● FABRIC_VPCEP_POLICY：网络打通功能相关权限策略。 ● FABRIC_DNS_POLICY：DNS相关权限策略。 ● FABRIC_OBS_POLICY：OBS相关权限策略。 默认取值：不涉及。

请求参数

表 4-57 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

响应参数

状态码：200

表 4-58 响应 Body 参数

参数	参数类型	描述
-	AnyType	OK

状态码：400

表 4-59 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释： 错误码。 约束限制： 不涉及。 取值范围： 长度为[8,36]个字符。 默认取值： 不涉及。
error_msg	String	参数解释： 错误描述。 约束限制： 不涉及。 取值范围： 长度为[2,512]个字符。 默认取值： 不涉及。
solution_msg	String	参数解释： 解决方案描述。 约束限制： 不涉及。 取值范围： 长度为[0,4096]个字符。 默认取值： 不涉及。

状态码：401

表 4-60 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-61 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-62 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-63 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

GET https://{endpoint}/v1/agency

响应示例

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	OK
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.4 工作空间管理

4.4.1 查询工作空间列表

功能介绍

列举工作空间列表。用户可通过该接口获取符合过滤条件的工作空间列表，接口支持分页查询、支持通过名称、ID和企业项目ID查询。返回为符合过滤条件的工作空间列表。此接口为同步接口，无配套使用接口和特殊场景。

URI

GET /v1/workspaces

表 4-64 Query 参数

参数	是否必选	参数类型	描述
offset	否	Integer	参数解释： 偏移量，表示从此偏移量开始查询。 约束限制： 不涉及。 取值范围： [0,10000000000]。 默认取值： 0。

参数	是否必选	参数类型	描述
limit	否	Integer	参数解释： 指定每一页返回的最大条目数。 约束限制： 不涉及。 取值范围： [1,100]。 默认取值： 10。
name	否	String	参数解释： 工作空间名称。 约束限制： 不涉及。 取值范围： 长度为[4,32]的字母、数字、中划线(-)和下划线(_)的组合。 默认取值： 不涉及。
id	否	String	参数解释： 工作空间ID，获取方法，请参见 获取工作空间ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
enterprise_project_id	是	String	参数解释： 企业项目ID。获取方法，请参见 获取企业项目ID 。 约束限制： 长度为[1,36]的英文字符、数字、中划线(-)和下划线(_)的组合。 取值范围： 不涉及。 默认取值： all_granted_eps。

请求参数

表 4-65 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

响应参数

状态码：200

表 4-66 响应 Body 参数

参数	参数类型	描述
total	Integer	参数解释：Workspace总数。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
workspaces	Array of Workspace objects	参数解释：Workspace列表。 约束限制：[0,1000]。

表 4-67 Workspace

参数	参数类型	描述
id	String	参数解释：工作空间ID。 约束限制：不涉及。 取值范围：长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值：不涉及。
name	String	参数解释：工作空间名称。 约束限制：不涉及。 取值范围：长度为[4,32]的英文字符、数字和中划线(-)的组合。 默认取值：不涉及。
description	String	参数解释：用户输入的描述。 约束限制：不涉及。 取值范围：长度为[0,255]个字符。 默认取值：不涉及。
create_time	String	参数解释：创建时间。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
create_domain_name	String	参数解释：创建账号名称。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

参数	参数类型	描述
create_user_name	String	参数解释： 创建用户名称。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
metastore_id	String	参数解释： LakeFormation服务的实例ID，即MetaStoreID。 约束限制： 不涉及。 取值范围： 长度为[32,36]的字母、数字、中划线(-)组合。 默认取值： 不涉及。
access_url	String	参数解释： 访问资源地址。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
enterprise_project_id	String	参数解释： 企业项目ID。获取方法，请参见 获取企业项目ID 。 约束限制： 只有对接了企业项目才可以填写。 取值范围： 长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值： 0，即default。
tags	Array of ResourceTag objects	参数解释： 工作空间的资源标签。 约束限制： 不涉及。

表 4-68 ResourceTag

参数	参数类型	描述
key	String	参数解释： 资源标签的键。 约束限制： 不涉及。 取值范围： 长度1到128的多语言字母、空格、制表符、数字、下划线(_)、点(.)、冒号(:)、斜杠(/)、等号(=)、加号(+)、减号(-)、@符号(@)组合，见 查询标签键列表 。 默认取值： 不涉及。

参数	参数类型	描述
value	String	参数解释：资源标签的值。 约束限制：不涉及。 取值范围：长度0到255的多语言字母、空格、制表符、数字、下划线(_)、点(.)、冒号(:)、斜杠(/)、等号(=)、加号(+)、减号(-)、@符号(@)组合。 默认取值：不涉及。

状态码：400

表 4-69 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-70 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404**表 4-71 响应 Body 参数**

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408**表 4-72 响应 Body 参数**

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-73 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

GET https://{endpoint}/v1/workspaces

响应示例

状态码：200

查询工作空间列表响应体。

```
{
  "total" : 2,
  "workspaces" : [ {
    "id" : "90a827a1-ca72-4be8-943b-774f94xxxxxx",
    "name" : "workspace1SqlResourceDef",
```

```
"create_time" : "2025-01-10T03:37:12.000+00:00",
"create_domain_name" : "create_domain_name_001",
"create_user_name" : "create_user_name_001",
"enterprise_project_id" : 0
}, {
  "id" : "90a827a1-ca72-4be8-943b-774f95xxxxx",
  "name" : "workspace2",
  "create_time" : "2025-01-11T03:37:12.000+00:00",
  "create_domain_name" : "create_domain_name_002",
  "create_user_name" : "create_user_name_002",
  "enterprise_project_id" : 0
}]
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	查询工作空间列表响应体。
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.4.2 更新工作空间

功能介绍

更新工作空间。用户可通过该接口更新工作空间，支持传入名称、描述、MetaStoreID，返回更新后的工作空间详情。此接口为同步接口，无配套使用接口和特殊场景。

URI

PUT /v1/workspaces/{workspace_id}

表 4-74 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间ID。获取方法，请参见 获取工作空间ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-75 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

表 4-76 请求 Body 参数

参数	是否必选	参数类型	描述
name	否	String	参数解释： 工作空间名称。 约束限制： 不涉及。 取值范围： 长度为[4,32]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
description	否	String	参数解释： 用户输入的描述。 约束限制： 不涉及。 取值范围： 长度为[0,255]个字符。 默认取值： 不涉及。
metastore_id	否	String	参数解释： LakeFormation服务的实例ID，即MetaStoreID。 约束限制： 不涉及。 取值范围： 长度为[32,36]的字母、数字、中划线(-)组合。 默认取值： 不涉及。
enterprise_project_id	否	String	参数解释： 企业项目ID。获取方法，请参见 获取企业项目ID 。 约束限制： 只有对接了企业项目才可以填写。 取值范围： 长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值： 0，即default。

响应参数

状态码：200

表 4-77 响应 Body 参数

参数	参数类型	描述
id	String	参数解释：工作空间ID。 约束限制：不涉及。 取值范围：长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值：不涉及。
name	String	参数解释：工作空间名称。 约束限制：不涉及。 取值范围：长度为[4,32]的英文字符、数字和中划线(-)的组合。 默认取值：不涉及。
description	String	参数解释：用户输入的描述。 约束限制：不涉及。 取值范围：长度为[0,255]个字符。 默认取值：不涉及。
create_time	String	参数解释：创建时间。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
create_domain_name	String	参数解释：创建账号名称。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
create_user_name	String	参数解释：创建用户名称。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
metastore_id	String	参数解释：LakeFormation服务的实例ID，即MetaStoreID。 约束限制：不涉及。 取值范围：长度为[32,36]的字母、数字、中划线(-)组合。 默认取值：不涉及。

参数	参数类型	描述
access_url	String	参数解释： 访问资源地址。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
enterprise_project_id	String	参数解释： 企业项目ID。获取方法，请参见 获取企业项目ID 。 约束限制： 只有对接了企业项目才可以填写。 取值范围： 长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值： 0，即default。
tags	Array of ResourceTag objects	参数解释： 工作空间的资源标签。 约束限制： 不涉及。

表 4-78 ResourceTag

参数	参数类型	描述
key	String	参数解释： 资源标签的键。 约束限制： 不涉及。 取值范围： 长度1到128的多语言字母、空格、制表符、数字、下划线(_)、点(.)、冒号(:)、斜杠(/)、等号(=)、加号(+)、减号(-)、@符号(@)组合，见 查询标签键列表 。 默认取值： 不涉及。
value	String	参数解释： 资源标签的值。 约束限制： 不涉及。 取值范围： 长度0到255的多语言字母、空格、制表符、数字、下划线(_)、点(.)、冒号(:)、斜杠(/)、等号(=)、加号(+)、减号(-)、@符号(@)组合。 默认取值： 不涉及。

状态码：400

表 4-79 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-80 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-81 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-82 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-83 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

更新工作空间名称为fabric，更新描述为fabric-description，更新对应MetastoreID为XXXXXX。

```
PUT https://{endpoint}/v1/workspaces/{workspace_id}
{
  "name": "fabric",
  "description": "fabric-description",
  "metastore_id": "xxxxxx"
}
```

响应示例

状态码：200

更新工作空间列表响应体。

```
{
  "id": "exxxx0ef-f4eb-4b95-aff1-9d33ae9fxxxx",
  "name": "fabric",
  "description": "fabric-description",
  "create_time": "2023-05-30T12:24:30.401Z",
  "create_domain_name": "domain_name",
  "create_user_name": "user_name",
  "metastore_id": "2xxx518f-42b8-4947-b20b-adfcxxx81a2x",
  "access_url": "https://test.fabric.com/",
  "enterprise_project_id": "01049549-82cd-4b2b-9733-ddb9435xxxx"
}
```

状态码：400

BadRequest

```
{
  "error_code": "common.01000001",
  "error_msg": "failed to read http request, please check your input, code: 400, reason: Type mismatch,"
}
```

```
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	更新工作空间列表响应体。
400	BadRequest
401	Unauthorized
403	Forbidden

状态码	描述
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.4.3 创建工作空间

功能介绍

创建工作空间。用户可通过该接口创建工作空间，支持传入名称、描述、MetaStoreID、企业项目ID和标签，返回工作空间详情。此接口为同步接口，无配套使用接口。创建工作空间接口存在同名校验，不支持创建重复名称的工作空间，并且同一账户下创建工作空间数量存在上限。

URI

POST /v1/workspaces

请求参数

表 4-84 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户Token。用于调用接口的认证。获取方法，请参见认证鉴权。 约束限制：不涉及。 取值范围：长度不超过65534个字符。 默认取值：不涉及。

表 4-85 请求 Body 参数

参数	是否必选	参数类型	描述
name	是	String	参数解释： 工作空间名称。 约束限制： 不涉及。 取值范围： 长度为[4,32]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
description	否	String	参数解释： 用户输入的描述。 约束限制： 不涉及。 取值范围： 长度为[0,255]个字符。 默认取值： 不涉及。
metastore_id	否	String	参数解释： LakeFormation服务的实例ID，即MetaStoreID。 约束限制： 不涉及。 取值范围： 长度为[32,36]的字母、数字、中划线(-)组合。 默认取值： 不涉及。
enterprise_project_id	否	String	参数解释： 企业项目ID。获取方法，请参见 获取企业项目ID 。 约束限制： 只有对接了企业项目才可以填写。 取值范围： 长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值： 0，即default。
tags	否	Array of ResourceTag objects	参数解释： 资源标签。 约束限制： 不涉及。

表 4-86 ResourceTag

参数	是否必选	参数类型	描述
key	是	String	参数解释：资源标签的键。 约束限制：不涉及。 取值范围：长度1到128的多语言字母、空格、制表符、数字、下划线(_)、点(.)、冒号(:)、斜杠(/)、等号(=)、加号(+)、减号(-)、@符号(@)组合，见查询标签键列表。 默认取值：不涉及。
value	否	String	参数解释：资源标签的值。 约束限制：不涉及。 取值范围：长度0到255的多语言字母、空格、制表符、数字、下划线(_)、点(.)、冒号(:)、斜杠(/)、等号(=)、加号(+)、减号(-)、@符号(@)组合。 默认取值：不涉及。

响应参数

状态码：200

表 4-87 响应 Body 参数

参数	参数类型	描述
id	String	参数解释：工作空间ID。 约束限制：不涉及。 取值范围：长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值：不涉及。
name	String	参数解释：工作空间名称。 约束限制：不涉及。 取值范围：长度为[4,32]的英文字符、数字和中划线(-)的组合。 默认取值：不涉及。
description	String	参数解释：用户输入的描述。 约束限制：不涉及。 取值范围：长度为[0,255]个字符。 默认取值：不涉及。

参数	参数类型	描述
create_time	String	参数解释：创建时间。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
create_domain_name	String	参数解释：创建账号名称。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
create_user_name	String	参数解释：创建用户名称。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
metastore_id	String	参数解释：LakeFormation服务的实例ID，即MetaStoreID。 约束限制：不涉及。 取值范围：长度为[32,36]的字母、数字、中划线(-)组合。 默认取值：不涉及。
access_url	String	参数解释：访问资源地址。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
enterprise_project_id	String	参数解释：企业项目ID。获取方法，请参见 获取企业项目ID 。 约束限制：只有对接了企业项目才可以填写。 取值范围：长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值：0，即default。
tags	Array of ResourceTag objects	参数解释：工作空间的资源标签。 约束限制：不涉及。

表 4-88 ResourceTag

参数	参数类型	描述
key	String	参数解释：资源标签的键。 约束限制：不涉及。 取值范围：长度1到128的多语言字母、空格、制表符、数字、下划线(_)、点(.)、冒号(:)、斜杠(/)、等号(=)、加号(+)、减号(-)、@符号(@)组合，见查询标签键列表。 默认取值：不涉及。
value	String	参数解释：资源标签的值。 约束限制：不涉及。 取值范围：长度0到255的多语言字母、空格、制表符、数字、下划线(_)、点(.)、冒号(:)、斜杠(/)、等号(=)、加号(+)、减号(-)、@符号(@)组合。 默认取值：不涉及。

状态码：400

表 4-89 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-90 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-91 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-92 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-93 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

创建工作空间。其名称为fabric，描述为fabric-description，对应的MetastoreID为xxxxxx，对应企业项目ID为01049549-82cd-4b2b-9733-ddb94350c125，对应标签有key1，值为value1，key2，值为value2。

POST https://{endpoint}/v1/workspaces

```
{
  "name" : "fabric",
  "description" : "fabric-description",
  "metastore_id" : "xxxxxx",
  "enterprise_project_id" : "01049549-82cd-4b2b-9733-ddb94350c125",
  "tags" : [ {
    "key" : "key1",
    "value" : "value1"
  }, {
    "key" : "key2",
    "value" : "value2"
  } ]
}
```

响应示例

状态码：200

创建工作空间列表响应体。

```
{
  "id" : "90a827a1-ca72-4be8-943b-774f94xxxxxx",
  "name" : "workspace1",
  "description" : "workspace1-description",
  "create_time" : "2025-01-10T03:37:12.000+00:00",
  "create_domain_name" : "create_domain_name_001",
  "create_user_name" : "create_user_name_001",
  "metastore_id" : "2180518f-42b8-4947-b20b-adfc539xxxxx",
  "access_url" : "https://test.fabric.com/",
  "enterprise_project_id" : 0
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch., cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	创建工作空间列表响应体。
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.4.4 删除工作空间

功能介绍

删除工作空间。此接口为同步接口，无配套使用接口。在该工作空间下存在计算资源、端点以及App的情况下无法删除，需要先清除该工作空间下创建的各种资源后才能删除。

URI

DELETE /v1/workspaces/{workspace_id}

表 4-94 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： workspaceID。获取方法，请参见 获取workspaceID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-95 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

响应参数

状态码：200

删除workspace响应体。

状态码：400

表 4-96 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释： 错误码。 约束限制： 不涉及。 取值范围： 长度为[8,36]个字符。 默认取值： 不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-97 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-98 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-99 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-100 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

```
DELETE https://{endpoint}/v1/workspaces/{workspace_id}
```

响应示例

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error": {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	删除工作空间响应体。
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.5 服务规格

4.5.1 查询资源规格列表

功能介绍

列举资源规格列表。用户可通过该接口列举资源规格列表，用户可以通过编码、类型、使用场景等参数过滤查询，支持分页。此接口为同步接口，无配套使用接口和特殊场景。

URI

GET /v1/specs

表 4-101 Query 参数

参数	是否必选	参数类型	描述
spec_code	否	String	规格编码。
spec_types	否	Array of strings	参数解释： 通过资源规格类型检索集合。 约束限制： 从规格列表查询获取。 取值范围： [0,100]。 默认取值： 不涉及。
offset	否	Integer	参数解释： 偏移量，表示从此偏移量开始查询。 约束限制： 不涉及。 取值范围： [0,1000000000]。 默认取值： 0。
limit	否	Integer	参数解释： 指定每一页返回的最大条目数。 约束限制： 不涉及。 取值范围： [1,100]。 默认取值： 10。
scenario	否	Object	参数解释： 规格使用场景。 约束限制： 不涉及。 取值范围： ● 不填：表示不限制； ● COMPUTE: 用于购买Ray计算资源时配置的物理节点规格； ● ENDPOINT: 用于创建Endpoint时配置的资源组规格大小。 默认取值： 不涉及。

请求参数

无

响应参数

状态码：200

表 4-102 响应 Body 参数

参数	参数类型	描述
total	Integer	参数解释：总数。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
spec_codes	Array of Spec objects	参数解释：规格编码列表。 约束限制：不涉及。 取值范围：[0,100]。 默认取值：不涉及。

表 4-103 Spec

参数	参数类型	描述
spec_code	String	参数解释：资源规格。 约束限制：不涉及。 取值范围：从规格列表查询获取。 默认取值：不涉及。
spec_type	String	参数解释：资源规格类型。 约束限制：不涉及。 取值范围： • APU：APU类型； • SQL_DPU：SQL DPU类型； • MU：MU类型； • RESOURCE_USAGE_UNIT：资源使用量类型； • RAY_APU：Ray APU类型； • RAY_DPU：Ray DPU类型。 • RAY_CACHE：Ray缓存资源。 默认取值：不涉及。
resource_type	String	参数解释：资源类型。 约束限制：不涉及。 取值范围：最小长度为1，最大长度为64。 默认取值：不涉及。

参数	参数类型	描述
stride	Integer	参数解释： 资源步长。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
unit	String	参数解释： 资源单位。 约束限制： 不涉及。 取值范围： 最小长度为1，最大长度为16。 默认取值： 不涉及。
min_stride_num	Integer	参数解释： 资源步数。 约束限制： 不涉及。 取值范围： 最小为1，最大为10000。输入范围还需要满足规格列表接口的步长白名单。 默认取值： 不涉及。
max_stride_num	Integer	参数解释： 资源步数。 约束限制： 不涉及。 取值范围： 最小为1，最大为10000。输入范围还需要满足规格列表接口的步长白名单。 默认取值： 不涉及。
usage_measure_id	Integer	参数解释： 使用量单位标识。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
usage_factor	String	参数解释： 使用量因子。 约束限制： 不涉及。 取值范围： 最小长度为1，最大长度为32。 默认取值： 不涉及。
stride_num_whitelist	Array of integers	参数解释： 步数白名单，返回时，步数必须是白名单中的值。 约束限制： 不涉及。 取值范围： 最小值为0，最大值为100。 默认取值： 不涉及。
extra_info	SpecExtraInfo object	参数解释： 附加信息。 约束限制： 不涉及。

表 4-104 SpecExtraInfo

参数	参数类型	描述
mu_multiple	Double	参数解释：模型单元换算倍数 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
mu_code	String	参数解释：模型单元编码。 约束限制：不涉及。 取值范围：长度[1,64]。 默认取值：不涉及。

状态码：400

表 4-105 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-106 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-107 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-108 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-109 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

GET https://{endpoint}/v1/specs

响应示例

状态码：200

查询规格列表响应。

```
{
  "total": 1,
  "spec_codes": [ {
    "spec_code": "string",
    "spec_type": "APU",
```

```
"resource_type" : "hws.resource.type.fabric.ray.dpu",
"stride" : 1,
"unit" : "string",
"min_stride_num" : 0,
"max_stride_num" : 1,
"usage_measure_id" : 6,
"usage_factor" : "string",
"stride_num_whitelist" : [ 12, 13 ],
"extra_info" : {
  "mu_multiple" : 0,
  "mu_code" : "string"
}
}]
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	查询规格列表响应。
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.6 端点管理

4.6.1 查询端点列表

功能介绍

列举端点列表。用户可使用该接口列举工作空间下端点列表，支持输入模型id、端点id、名称、类型、可见性等参数过滤，支持分页查询。此接口为同步接口，无配套使用接口和特殊场景。

URI

GET /v1/workspaces/{workspace_id}/endpoints

表 4-110 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释：工作空间ID，获取方法，请参见获取工作空间ID。 约束限制：不涉及。 取值范围：长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值：不涉及。

表 4-111 Query 参数

参数	是否必选	参数类型	描述
offset	否	Integer	参数解释：偏移量，表示从此偏移量开始查询。 约束限制：不涉及。 取值范围：[0,1000000000]。 默认取值：0。
limit	否	Integer	参数解释：指定每一页返回的最大条目数。 约束限制：不涉及。 取值范围：[1,100]。 默认取值：10。
model_id	否	String	参数解释：通过模型ID检索。 约束限制：不涉及。 取值范围：长度为[32,36]的英文字符、数字和中划线(-)的组合，请参见获取模型ID。 默认取值：不涉及。
name	否	String	参数解释：Endpoint名称，通过名字搜索Endpoint的参数。 约束限制：不涉及。 取值范围：长度为[1,128]的中文、字母、数字、下划线(_)、中划线(-)，支持模糊查询。 默认取值：不涉及。

参数	是否必选	参数类型	描述
endpoint_id	否	String	参数解释：通过端点ID检索Endpoint的参数。获取方法，请参见获取端点ID。 约束限制：不涉及。 取值范围：长度为[32,36]的英文字符、数字和中划线(-)的组合。 默认取值：不涉及。
type	否	Object	参数解释：Endpoint的类型。 约束限制：不涉及。 取值范围： • service: Service EP, 代表一个可接收Service请求资源组； • ray: RayCluter on k8s的EP, 代表一个Ray集群； • inference: 推理的EP, 代表一个推理函数实例； • job: Job EP, 代表一个可接收Job请求资源组； • ray_service: RayService on k8s的EP, 代表一个RayService。 • sql: SQL EP, 代表一个可接收SQL请求资源组。 默认取值：不涉及。
visibility	否	String	参数解释：可见性检索的参数。 约束限制：不涉及。 取值范围： • PRIVATE: 私有，用户自己创建的； • PUBLIC: 公共，查询所有公共的，包括其他用户创建的； • ALL: 所有的。 默认取值：默认为空，不填表示不限制，则查出当前用户下的，包括PRIVATE和PUBLIC，不包括其他用户创建的。

请求参数

表 4-112 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户Token。用于调用接口的认证。获取方法，请参见认证鉴权。 约束限制：不涉及。 取值范围：长度不超过65534个字符。 默认取值：不涉及。

响应参数

状态码：200

表 4-113 响应 Body 参数

参数	参数类型	描述
total	Integer	参数解释：符合条件的端点总数。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
endpoints	Array of EndpointBriefInfo objects	参数解释：符合条件的Endpoint简要信息列表。 约束限制： [0,1000]。

表 4-114 EndpointBriefInfo

参数	参数类型	描述
visibility	String	参数解释：可见性。 约束限制：不涉及。 取值范围： - PRIVATE：私有 - PUBLIC：公共 默认取值： PRIVATE。

参数	参数类型	描述
id	String	参数解释：Endpoint ID。 约束限制：不涉及。 取值范围：长度为[32,36]的英文字符、数字和中划线(-)的组合。 默认取值：不涉及。
name	String	参数解释：Endpoint名称。 约束限制：不涉及。 取值范围：长度为[1,128]包含中文、字母、数字、下划线、中划线、半角句号(.)、空格的组合。 默认取值：不涉及。
type	String	参数解释：Endpoint的类型。 约束限制：不涉及。 取值范围： • service: Service EP, 代表一个可接收Service请求资源组; • ray: RayCluter on k8s的EP, 代表一个Ray集群; • inference: 推理的EP, 代表一个推理函数实例; • job: Job EP, 代表一个可接收Job请求资源组; • ray_service: RayService on k8s的EP, 代表一个RayService。 • sql: SQL EP, 代表一个可接收SQL请求资源组。 默认取值：不涉及。

参数	参数类型	描述
status	String	参数解释：Endpoint状态。 约束限制：不涉及。 取值范围： • CREATING：创建中； • RUNNING：运行中； • CREATE_FAILED：创建失败； • DELETING：删除中； • DELETED：已删除； • UPDATING：更新中； • UPDATE_FAILED：升级失败； • DELETE_FAILED：创建失败； • FROZEN：冻结； • INACTIVE：未开通（公共EP）； • TRIAL_ENDED：试用结束。 默认取值：不涉及。
description	String	参数解释：描述信息。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
create_time	String	参数解释：创建时间。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
update_time	String	参数解释：更新时间。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
owner	User object	参数解释：租户信息内容，包含账户名称、账户ID和用户名称、用户ID。 约束限制：不涉及。
cap	CapRef object	参数解释：镜像包信息。 约束限制：不涉及。
reserved_resource	ReservedResource object	参数解释：通用预留资源。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

参数	参数类型	描述
ray_resource	RayResourceInfo object	参数解释：Ray资源配置。 约束限制：不涉及。
caches	Array of CacheConfig objects	参数解释：缓存关联信息。 约束限制：不涉及。 取值范围：[0,1000]。 默认取值：不涉及。
error_code	String	参数解释：失败状态时的错误编码。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
error_msg	String	参数解释：失败状态时的错误信息。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
public_access	Boolean	参数解释：是否开启公网访问。 约束限制：不涉及。 取值范围：开启true，关闭false。 默认取值：不涉及。

表 4-115 User

参数	参数类型	描述
domain_name	String	参数解释：账户名称。 约束限制：不涉及。 取值范围：长度为[0,64]个字符。 默认取值：不涉及。
domain_id	String	参数解释：账号ID。 约束限制：不涉及。 取值范围：长度为[1,36]的数字、英文字符和中划线(-)的组合。 默认取值：不涉及。
user_name	String	参数解释：用户名称。 约束限制：不涉及。 取值范围：长度为[0,64]个字符。 默认取值：不涉及。

参数	参数类型	描述
user_id	String	参数解释： 用户ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的数字、英文字符和中划线(-)的组合。 默认取值： 不涉及。

表 4-116 CapRef

参数	参数类型	描述
cap_id	String	参数解释： 镜像包ID。 约束限制： 不涉及。 取值范围： 长度为[32,36]的英文、数字、中划线(-)的组合。 默认取值： 不涉及。
version_id	String	参数解释： 镜像包版本ID。 约束限制： 不涉及。 取值范围： 长度为[32,36]的英文、数字、中划线(-)的组合。 默认取值： 不涉及。

表 4-117 ReservedResource

参数	参数类型	描述
apu	ResourceDef object	参数解释： 资源定义信息。 约束限制： 不涉及。
dpu	SqlResourceDef object	参数解释： Sql资源定义信息。 约束限制： 不涉及。
mu	ResourceDemand object	参数解释： 资源需求量配置。 约束限制： 不涉及。

表 4-118 ResourceDef

参数	参数类型	描述
spec_code	String	参数解释： Endpoint状态。 约束限制： 资源规格编码，从查询规格列表ListSpecs接口获取。 取值范围： 不涉及。 默认取值： 不涉及。
warm_up_num	Integer	参数解释： 预热资源量。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
max_num	Integer	参数解释： 最大资源量，不填默认为预热资源量，即不使用弹性资源。 约束限制： 不涉及。 取值范围： 最小值为1，最大值为1000。 默认取值： 不涉及。

表 4-119 SqlResourceDef

参数	参数类型	描述
spec_code	String	参数解释： Endpoint状态。 约束限制： 资源规格编码，从查询规格列表ListSpecs接口获取。 取值范围： 不涉及。 默认取值： 不涉及。
warm_up_num	Integer	参数解释： 预热资源量。 约束限制： 不涉及。 取值范围： 最小值为1，最大值为5000。 默认取值： 不涉及。
max_num	Integer	参数解释： 最大资源量，不填默认为预热资源量，即不使用弹性资源。 约束限制： 不涉及。 取值范围： 最小值为1，最大值为5000。 默认取值： 不涉及。
use_public_resource	Boolean	参数解释： 是否使用公共资源。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

表 4-120 ResourceDemand

参数	参数类型	描述
spec_code	String	参数解释：资源规格。 约束限制：不涉及。 取值范围：从规格列表查询获取。 默认取值：不涉及。
min	Integer	参数解释：最小数。 约束限制：不涉及。 取值范围：最小值为0。 默认取值：不涉及。
max	Integer	参数解释：最大数。 约束限制：不涉及。 取值范围：最小值为1。 默认取值：不涉及。

表 4-121 RayResourceInfo

参数	参数类型	描述
head_group_spec	HeadGroupSpec object	参数解释：Ray Head配置。 约束限制：不涉及。
worker_group_spec	Array of WorkerGroupSpec objects	参数解释：Ray worker group配置。 约束限制：不涉及。 取值范围：[1,10]。 默认取值：不涉及。

表 4-122 HeadGroupSpec

参数	参数类型	描述
spec_code	String	参数解释：资源规格。 约束限制：不涉及。 取值范围：从规格列表查询获取。 默认取值：不涉及。

表 4-123 WorkerGroupSpec

参数	参数类型	描述
name	String	参数解释：名称。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
spec_code	String	参数解释：资源规格。 约束限制：不涉及。 取值范围：从规格列表查询获取。 默认取值：不涉及。
min_replicas	Integer	参数解释：最小副本数。 约束限制：不涉及。 取值范围：最小值为0，最大值为10000。 默认取值：不涉及。
max_replicas	Integer	参数解释：最大副本数。 约束限制：不涉及。 取值范围：最小值为1，最大值为10000。 默认取值：不涉及。

表 4-124 CacheConfig

参数	参数类型	描述
mnt_path	String	参数解释：挂载路径。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
cache_ins_id	String	参数解释：分布式缓存id。 约束限制：不涉及。 取值范围：已创建的分布式缓存资源id。 默认取值：不涉及。

状态码：400

表 4-125 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-126 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-127 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-128 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

```
GET https://{endpoint}/v1/workspaces/{workspace_id}/endpoints
```

响应示例

状态码：200

查询Endpoint列表的响应体。

```
{
  "total": 1,
  "endpoints": [ {
    "visibility": "PRIVATE",
    "id": "0b5633ba2b904511ad514346f4dxxxxx",
    "name": "endpoint1",
    "type": "job",
    "status": "CREATING",
    "description": "description",
    "create_time": "2023-05-30T12:24:30.401Z",
    "update_time": "2023-05-30T12:24:30.401Z",
    "owner": {
      "domain_name": "string",
      "domain_id": "0b5633ba2b904511ad514346f4dxxxxx",
      "user_name": "string",
      "user_id": "0b5633ba2b904511ad514346f4dxxxxx"
    },
    "cap": {
      "cap_id": "0b5633ba2b904511ad514346f4dxxxxx",
      "version_id": "0b5633ba2b904511ad514346f4dxxxxx"
    },
    "reserved_resource": {
      "apu": {
        "spec_code": "a1x",
        "warm_up_num": 1,
        "max_num": 100
      },
      "dpu": {
        "spec_code": "d1x",
        "warm_up_num": 1,
        "max_num": 100,
        "use_public_resource": false
      },
      "mu": {
        "spec_code": "mu",
        "min": 0,
        "max": 1
      }
    },
    "ray_resource": {
      "head_group_spec": {
        "spec_code": "string"
      },
      "worker_group_spec": [ {
        "name": "string",
        "spec_code": "string",
        "min_replicas": 10000,
        "max_replicas": 10000
      } ]
    },
    "caches": [ {
      "mnt_path": "string",
      "cache_ins_id": "string"
    } ],
    "error_code": "string",
    "error_msg": "string"
  } ]
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch., cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	查询Endpoint列表的响应体。
400	BadRequest
401	Unauthorized
403	Forbidden
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.6.2 创建端点

功能介绍

创建端点。用户可使用该接口在工作空间下创建端点，输入名称、描述、类型、资源配置、cap配置等信息，返回端点信息。此接口为异步接口，配套使用接口ShowEndpoint查询端点详情，无特殊场景。

URI

POST /v1/workspaces/{workspace_id}/endpoints

表 4-129 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间ID，获取方法，请参见 获取工作空间ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-130 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

表 4-131 请求 Body 参数

参数	是否必选	参数类型	描述
name	是	String	参数解释： Endpoint名称。 约束限制： 不涉及。 取值范围： 长度为[1,128]包含中文、字母、数字、下划线、中划线、半角句号(.)、空格的组合。 默认取值： 不涉及。
description	否	String	参数解释： 描述信息。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
type	是	String	参数解释： Endpoint的类型。 约束限制： 不涉及。 取值范围： • service: Service EP, 代表一个可接收Service请求资源组; • ray: RayCluter on k8s的EP, 代表一个Ray集群; • inference: 推理的EP, 代表一个推理函数实例; • job: Job EP, 代表一个可接收Job请求资源组; • ray_service: RayService on k8s的EP, 代表一个RayService。 • sql: SQL EP, 代表一个可接收SQL请求资源组。 默认取值： 不涉及。
reserved_resource	否	ReservedResource object	参数解释： 通用预留资源。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
ray_resource	否	RayResourceInput object	参数解释： 创建Ray集群的资源配置。 约束限制： 不涉及。
cap	否	CapRef object	参数解释： 镜像包信息。 约束限制： 不涉及。

参数	是否必选	参数类型	描述
config	否	EndpointConfig object	参数解释：端点的配置。 约束限制：不涉及。
cache_list	否	Array of CacheConfig objects	参数解释：分布式缓存配置。 约束限制：不涉及。
public_access	否	Boolean	参数解释：是否开启公网访问。 约束限制：不涉及。 取值范围：开启true，关闭false。 默认取值：false。

表 4-132 ReservedResource

参数	是否必选	参数类型	描述
apu	否	ResourceDef object	参数解释：资源定义信息。 约束限制：不涉及。
dpu	否	SqlResourceDef object	参数解释：Sql资源定义信息。 约束限制：不涉及。
mu	否	ResourceDemand object	参数解释：资源需求量配置。 约束限制：不涉及。

表 4-133 ResourceDef

参数	是否必选	参数类型	描述
spec_code	是	String	参数解释：Endpoint状态。 约束限制：资源规格编码，从查询规格列表ListSpecs接口获取。 取值范围：不涉及。 默认取值：不涉及。
warm_up_num	是	Integer	参数解释：预热资源量。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

参数	是否必选	参数类型	描述
max_num	否	Integer	参数解释： 最大资源量，不填默认为预热资源量，即不使用弹性资源。 约束限制： 不涉及。 取值范围： 最小值为1，最大值为1000。 默认取值： 不涉及。

表 4-134 SqlResourceDef

参数	是否必选	参数类型	描述
spec_code	是	String	参数解释： Endpoint状态。 约束限制： 资源规格编码，从查询规格列表ListSpecs接口获取。 取值范围： 不涉及。 默认取值： 不涉及。
warm_up_num	是	Integer	参数解释： 预热资源量。 约束限制： 不涉及。 取值范围： 最小值为1，最大值为5000。 默认取值： 不涉及。
max_num	否	Integer	参数解释： 最大资源量，不填默认为预热资源量，即不使用弹性资源。 约束限制： 不涉及。 取值范围： 最小值为1，最大值为5000。 默认取值： 不涉及。
use_public_resource	否	Boolean	参数解释： 是否使用公共资源。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

表 4-135 RayResourceInput

参数	是否必选	参数类型	描述
head_node_resource	是	HeadNodeResourceDemand object	参数解释：headNode资源需求量配置。 约束限制：不涉及。
work_node_resources	是	Array of WorkNodeResourceDemand objects	参数解释：工作节点资源配置。 约束限制：不涉及。 取值范围：[1,10]。 默认取值：不涉及。

表 4-136 HeadNodeResourceDemand

参数	是否必选	参数类型	描述
spec_code	是	String	参数解释：资源规格。 约束限制：不涉及。 取值范围：从规格列表查询获取。 默认取值：不涉及。

表 4-137 WorkNodeResourceDemand

参数	是否必选	参数类型	描述
name	是	String	参数解释：资源配置名称。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
resource	是	ResourceDemand object	参数解释：资源需求量配置。 约束限制：不涉及。

表 4-138 ResourceDemand

参数	是否必选	参数类型	描述
spec_code	是	String	参数解释：资源规格。 约束限制：不涉及。 取值范围：从规格列表查询获取。 默认取值：不涉及。

参数	是否必选	参数类型	描述
min	是	Integer	参数解释：最小数。 约束限制：不涉及。 取值范围：最小值为0。 默认取值：不涉及。
max	是	Integer	参数解释：最大数。 约束限制：不涉及。 取值范围：最小值为1。 默认取值：不涉及。

表 4-139 CapRef

参数	是否必选	参数类型	描述
cap_id	否	String	参数解释：镜像包ID。 约束限制：不涉及。 取值范围：长度为[32,36]的英文、数字、中划线(-)的组合。 默认取值：不涉及。
version_id	否	String	参数解释：镜像包版本ID。 约束限制：不涉及。 取值范围：长度为[32,36]的英文、数字、中划线(-)的组合。 默认取值：不涉及。

表 4-140 EndpointConfig

参数	是否必选	参数类型	描述
ray_service_config	否	RayServiceConfig object	参数解释：RayService类型端点的配置。 约束限制：不涉及。

表 4-141 RayServiceConfig

参数	是否必选	参数类型	描述
ray_serve_config	是	RayServeConfig object	参数解释：RayServe配置。 约束限制：不涉及。

参数	是否必选	参数类型	描述
ray_cluster_config	是	RayClusterConfig object	参数解释：Ray集群配置。 约束限制：不涉及。
log_config	否	RayServiceLogConfig object	参数解释：RayService日志配置。 约束限制：不涉及。
data_infos	否	Array of DataBriefInfo objects	参数解释：数据信息。 约束限制：不涉及。 取值范围：[0,1]。 默认取值：不涉及。

表 4-142 RayServeConfig

参数	是否必选	参数类型	描述
applications	否	Array of ServeApplication objects	参数解释：Application。 约束限制：不涉及。 取值范围：[0,5]。 默认取值：不涉及。

表 4-143 ServeApplication

参数	是否必选	参数类型	描述
name	是	String	参数解释：Application的名称。 约束限制：不涉及。 取值范围：长度为[1,64]的中文、字母、数字、下划线、中划线、半角句号(.)、空格的组合。 默认取值：不涉及。
import_path	是	String	参数解释：输入路径。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
deployments	是	Array of Deployment objects	参数解释：Deployment列表。 约束限制：不涉及。 取值范围：[1,5]。 默认取值：不涉及。

参数	是否必选	参数类型	描述
runtime_env	是	ServeRuntimeEnv object	参数解释： Serve的运行时环境配置。 约束限制： 不涉及。 取值范围： 可选参数有： working_dir: 代码将在其中运行的工作目录。必须是远程URI，如s3或本地路径;env_vars: 要设置的环境变量。 默认取值： 不涉及。
route_prefix	否	String	参数解释： 路由前缀。 约束限制： 要求在RayServe中具有唯一性，不能跟其他的Application的前缀重复。 取值范围： 不涉及。 默认取值： 不涉及。

表 4-144 Deployment

参数	是否必选	参数类型	描述
name	否	String	参数解释： Deployment的名称。 约束限制： 不涉及。 取值范围： 长度为[1,64]的中文、字母、数字、下划线、中划线、半角句号(.)、空格的组合。 默认取值： 不涉及。
ray_actor_options	否	RayActorOptions object	参数解释： RayActor配置。 约束限制： 不涉及。
autoscaling_config	否	AutoscalingConfig object	参数解释： 自动扩缩的配置。 约束限制： 不涉及。
num_replicas	否	Integer	参数解释： 副本数量。 约束限制： 不涉及。 取值范围： [0,1000]。 默认取值： 不涉及。
user_config	否	Object	参数解释： 用户自定义配置。 约束限制： 不涉及。

参数	是否必选	参数类型	描述
max_replicas_per_node	否	Integer	参数解释： 每个节点允许的最大副本数。 约束限制： 不涉及。 取值范围： [1,100]。 默认取值： 1。
max_ongoing_requests	否	Integer	参数解释： 每个副本可接受的最大并发请求数。 约束限制： 不涉及。 取值范围： [1,100000]。 默认取值： 不涉及。
max_queued_requests	否	Integer	参数解释： deployment可接受的排队的最大请求数。 -1时表示无限制。 约束限制： 不涉及。 取值范围： [-1,100000]。 默认取值： 不涉及。

表 4-145 RayActorOptions

参数	是否必选	参数类型	描述
num_cpus	否	Double	参数解释： CPU数量。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
memory	否	Double	参数解释： 内存数量。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
resources	否	String	参数解释： 资源配置。 约束限制： 不涉及。 取值范围： 长度[0,1024]。 默认取值： 不涉及。

表 4-146 AutoscalingConfig

参数	是否必选	参数类型	描述
target_ongoing_requests	否	Integer	参数解释：每个副本单位时间能提供的平均请求数。 约束限制：不涉及。 取值范围：[1,1000]。 默认取值：不涉及。
min_replicas	否	Integer	参数解释：最小副本数。 约束限制：不涉及。 取值范围：[0,1000]。 默认取值：不涉及。
max_replicas	否	Integer	参数解释：最大副本数。 约束限制：不涉及。 取值范围：[0,1000]。 默认取值：不涉及。
initial_replicas	否	Integer	参数解释：初始副本数。 约束限制：不涉及。 取值范围：[0,1000]。 默认取值：不涉及。
upscale_delay_s	否	Integer	参数解释：扩容之前的等待时间。 约束限制：不涉及。 取值范围：[0,86400]。 默认取值：不涉及。
downscale_delay_s	否	Integer	参数解释：缩容之前的等待时间。 约束限制：不涉及。 取值范围：[0,86400]。 默认取值：不涉及。

表 4-147 ServeRuntimeEnv

参数	是否必选	参数类型	描述
working_dir	否	String	参数解释： 代码将在其中运行的工作目录。 约束限制： 必须是远程URI，如s3或本地路径。 取值范围： 不涉及。 默认取值： 不涉及。
env_vars	否	Map<String,String>	参数解释： 要设置的环境变量。 约束限制： 不涉及。

表 4-148 RayClusterConfig

参数	是否必选	参数类型	描述
enable_custom	否	Boolean	参数解释： 是否开启资源自定义。 约束限制： 不涉及。 取值范围： 开启true，关闭false。 默认取值： false。
head_group_spec	是	HeadGroupSpecV2 object	参数解释： HeadGroup的配置。 约束限制： 不涉及。
worker_group_specs	是	Array of WorkerGroupSpecV2 objects	参数解释： Worker Group的配置。 约束限制： [1,1000]。

表 4-149 HeadGroupSpecV2

参数	是否必选	参数类型	描述
limits	否	ResourceSpec object	参数解释： 资源配置。 约束限制： 不涉及。
requests	否	ResourceSpec object	参数解释： 资源配置。 约束限制： 不涉及。

表 4-150 WorkerGroupSpecV2

参数	是否必选	参数类型	描述
name	否	String	参数解释：名称。 约束限制：不涉及。 取值范围：长度[0,128]字母、数字、下划线(_)、中划线(-)的组合。 默认取值：不涉及。
min_replicas	否	Integer	参数解释：最小副本数。 约束限制：不涉及。 取值范围：[0,10000]。 默认取值：不涉及。
max_replicas	否	Integer	参数解释：最大副本数。 约束限制：不涉及。 取值范围：[1,10000]。 默认取值：不涉及。
limits	否	ResourceSpec object	参数解释：资源限制。 约束限制：不涉及。
requests	否	ResourceSpec object	参数解释：最小资源数。 约束限制：不涉及。

表 4-151 ResourceSpec

参数	是否必选	参数类型	描述
spec_code	否	String	参数解释：资源规格。 约束限制：不涉及。 取值范围：从规格列表查询获取。 默认取值：不涉及。
cpu_resource	否	Integer	参数解释：cpu核数量。 约束限制：不涉及。 取值范围：[0,192]。 默认取值：null。
memory_resource	否	Integer	参数解释：内存大小。 约束限制：不涉及。 取值范围：[0,1536]。 默认取值：null。

参数	是否必选	参数类型	描述
npu_resource	否	Integer	参数解释： 昇腾卡数量。 约束限制： 不涉及。 取值范围： [0,8]。 默认取值： null。

表 4-152 RayServiceLogConfig

参数	是否必选	参数类型	描述
lts	否	RayServiceLogLtsConfig object	参数解释： RayService日志配置。 约束限制： 不涉及。

表 4-153 RayServiceLogLtsConfig

参数	是否必选	参数类型	描述
enabled	否	Boolean	参数解释： 是否开启RayService日志转储到LTS。 约束限制： 不涉及。 取值范围： true, false。 默认取值： false。
log_group_id	否	String	参数解释： RayService日志转储到的LTS日志组id。 约束限制： 需要是已经存在的LTS日志组id。 取值范围： 不涉及。 默认取值： 不涉及。
log_stream_id	否	String	参数解释： RayService日志转储到的LTS日志流id。 约束限制： 需要是已经存在的LTS日志流id。 取值范围： 不涉及。 默认取值： 不涉及

表 4-154 DataBriefInfo

参数	是否必选	参数类型	描述
input_path	否	String	参数解释：地址。 约束限制：不涉及。 取值范围：长度[4,1024]。 默认取值：不涉及。

表 4-155 CacheConfig

参数	是否必选	参数类型	描述
mnt_path	是	String	参数解释：挂载路径。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
cache_ins_id	是	String	参数解释：分布式缓存id。 约束限制：不涉及。 取值范围：已创建的分布式缓存资源id。 默认取值：不涉及。

响应参数

状态码：202

查询endpoint详情的响应。

状态码：400

表 4-156 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-157 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-158 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-159 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

创建端点。具体规格如样例展示。

```
POST https://{endpoint}/v1/workspaces/{workspace_id}/endpoints
{
  "name" : "endpoint1",
  "description" : "description",
  "type" : "job",
  "reserved_resource" : {
    "apu" : {
      "spec_code" : "dpu.d1x",
      "warm_up_num" : 1,
      "max_num" : 100
    },
    "dpu" : {
      "spec_code" : "dpu.d1x",
      "warm_up_num" : 4,
      "max_num" : 100
    },
    "mu" : {
      "spec_code" : "string",
```

```
    "min" : 0,
    "max" : 1
  }
},
"ray_resource" : {
  "head_node_resource" : {
    "spec_code" : "string"
  },
  "work_node_resources" : [ {
    "name" : "cpuGroup1",
    "resource" : {
      "spec_code" : "string",
      "min" : 0,
      "max" : 1
    }
  }
],
"cap" : {
  "cap_id" : "0b5633ba2b904511ad514346f4d23d4b",
  "version_id" : "0b5633ba2b904511ad514346f4d23d4b"
},
"config" : {
  "ray_service_config" : {
    "log_config" : {
      "lts" : {
        "enabled" : true,
        "log_group_id" : "string",
        "log_stream_id" : "string"
      }
    },
    "ray_serve_config" : {
      "applications" : [ {
        "name" : "string",
        "import_path" : "string",
        "deployments" : [ {
          "name" : "string",
          "ray_actor_options" : {
            "num_cpus" : 0,
            "memory" : 0,
            "resources" : "string"
          }
        },
        "autoscaling_config" : {
          "target_ongoing_requests" : 1000,
          "max_ongoing_requests" : 1000,
          "min_replicas" : 1000,
          "max_replicas" : 1000,
          "initial_replicas" : 1000,
          "upscale_delay_s" : 86400,
          "downscale_delay_s" : 86400
        },
        "num_replicas" : 1000,
        "user_config" : { },
        "max_replicas_per_node" : 1
      } ],
      "runtime_env" : {
        "working_dir" : "string",
        "env_vars" : {
          "additionalProp1" : "string",
          "additionalProp2" : "string",
          "additionalProp3" : "string"
        }
      },
      "route_prefix" : "string"
    }
  },
  "ray_cluster_config" : {
    "head_group_spec" : {
      "limits" : {
        "spec_code" : "string"
      }
    }
  }
}
```

```
    },
    "requests": {
      "spec_code": "string"
    }
  },
  "worker_group_specs": [ {
    "name": "string",
    "min_replicas": 10000,
    "max_replicas": 10000,
    "limits": {
      "spec_code": "string"
    },
    "requests": {
      "spec_code": "string"
    }
  }
],
"data_infos": [ {
  "input_path": "OBS://bucket/dir/"
}
],
"cache_list": [ {
  "mnt_path": "string",
  "cache_ins_id": "string"
}
]
```

响应示例

状态码：202

查询endpoint详情的响应。

```
{
  "visibility": "PRIVATE",
  "id": "0b5633ba2b904511ad514346f4d23d4b",
  "name": "endpoint1",
  "type": "job",
  "status": "CREATING",
  "description": "description",
  "create_time": "2023-05-30T12:24:30.401Z",
  "update_time": "2023-05-30T12:24:30.401Z",
  "owner": {
    "domain_name": "string",
    "domain_id": "0b5633ba2b904511ad514346f4d23d4b",
    "user_name": "string",
    "user_id": "0b5633ba2b904511ad514346f4d23d4b"
  },
  "cap": {
    "cap_id": "0b5633ba2b904511ad514346f4d23d4b",
    "version_id": "0b5633ba2b904511ad514346f4d23d4b"
  },
  "reserved_resource": {
    "apu": {
      "spec_code": "dpu.d1x",
      "warm_up_num": 1,
      "max_num": 100
    },
    "dpu": {
      "spec_code": "dpu.d1x",
      "warm_up_num": 1,
      "max_num": 100,
      "use_public_resource": false
    },
    "mu": {
      "spec_code": "string",
      "min": 0,
      "max": 1
    }
  }
}
```

```
}
},
"ray_resource": {
  "head_group_spec": {
    "spec_code": "string"
  },
  "worker_group_spec": [ {
    "name": "string",
    "spec_code": "string",
    "min_replicas": 10000,
    "max_replicas": 10000
  } ]
},
"cache": [ {
  "mnt_path": "string",
  "cache_ins_id": "string"
} ],
"error_code": "string",
"error_msg": "string",
"urls": [ {
  "name": "string",
  "address": "string",
  "type": "PRIVATE",
  "token": "string"
} ],
"business_engine_instance_ids": [ "string" ],
"tokens_quota": {
  "total": 0,
  "used": 0,
  "expire_time": "2025-01-17T01:12:24.663Z"
},
"runtime_env_type": "RELEASE",
"config": {
  "ray_service_config": {
    "log_config": {
      "lts": {
        "enabled": true,
        "log_group_id": "string",
        "log_stream_id": "string"
      }
    }
  },
  "ray_serve_config": {
    "applications": [ {
      "name": "string",
      "import_path": "string",
      "deployments": [ {
        "name": "string",
        "ray_actor_options": {
          "num_cpus": 0,
          "memory": 0,
          "resources": "string"
        }
      } ],
      "autoscaling_config": {
        "target_ongoing_requests": 1000,
        "max_ongoing_requests": 1000,
        "min_replicas": 1000,
        "max_replicas": 1000,
        "initial_replicas": 1000,
        "upscale_delay_s": 86400,
        "downscale_delay_s": 86400
      },
      "num_replicas": 1000,
      "user_config": { },
      "max_replicas_per_node": ""
    } ],
    "runtime_env": {
      "working_dir": "string",
      "env_vars": {
        "additionalProp1": "string",
```

```
        "additionalProp2" : "string",
        "additionalProp3" : "string"
      }
    },
    "route_prefix" : "string"
  } ]
},
"ray_cluster_config" : {
  "head_group_spec" : {
    "limits" : {
      "spec_code" : "string"
    },
    "requests" : {
      "spec_code" : "string"
    }
  },
  "worker_group_specs" : [ {
    "name" : "string",
    "min_replicas" : 10000,
    "max_replicas" : 10000,
    "limits" : {
      "spec_code" : "string"
    },
    "requests" : {
      "spec_code" : "string"
    }
  } ]
},
"data_infos" : [ {
  "input_path" : "OBS://bucket/dir/",
  "output_path" : "OBS://bucket/dir/",
  "env_var_name" : "string"
} ]
}
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：408

Request Time-out

```
{
  "error_code": "common.00000408",
  "error_msg": "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code": "common.00000500",
  "error_msg": "internal error"
}
```

状态码

状态码	描述
202	查询endpoint详情的响应。
400	BadRequest
401	Unauthorized
403	Forbidden
408	Request Time-out
500	InternalServerError

错误码请参见[错误码](#)。**4.6.3 查询端点详情****功能介绍**

查询端点详情。用户可通过该接口查询某端点详情，输入为端点id，返回对应端点详情。此接口为同步接口，无配套使用接口和特殊场景。

URI

GET /v1/workspaces/{workspace_id}/endpoints/{endpoint_id}

表 4-160 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间ID，获取方法，请参见 获取工作空间ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
endpoint_id	是	String	参数解释： 端点ID。获取方法，请参见 获取端点ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

表 4-161 Query 参数

参数	是否必选	参数类型	描述
version	否	String	参数解释： 查询Endpoint信息的版本。 约束限制： 不涉及。 取值范围： • CURRENT：当前版本； • PREVIOUS：上个版本。 默认取值： 不涉及。

请求参数

表 4-162 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

响应参数

状态码：200

查询endpoint详情的响应。

状态码：400

表 4-163 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-164 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-165 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-166 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

GET https://{endpoint}/v1/workspaces/{workspace_id}/endpoints/{endpoint_id}

响应示例

状态码：200

查询endpoint详情的响应。

```
{
  "visibility": "PRIVATE",
  "id": "0b5633ba2b904511ad514346f4d23d4b",
  "name": "endpoint1",
  "type": "job",
  "status": "CREATING",
  "description": "description",
  "create_time": "2023-05-30T12:24:30.401Z",
  "update_time": "2023-05-30T12:24:30.401Z",
  "owner": {
    "domain_name": "string",
    "domain_id": "0b5633ba2b904511ad514346f4d23d4b",
    "user_name": "string",
    "user_id": "0b5633ba2b904511ad514346f4d23d4b"
  },
  "cap": {
    "cap_id": "0b5633ba2b904511ad514346f4d23d4b",
    "version_id": "0b5633ba2b904511ad514346f4d23d4b"
  },
  "reserved_resource": {
    "apu": {
      "spec_code": "dpu.d1x",
      "warm_up_num": 1,
      "max_num": 100
    },
    "dpu": {
      "spec_code": "dpu.d1x",
      "warm_up_num": 1,
      "max_num": 100,
      "use_public_resource": false
    },
    "mu": {
      "spec_code": "string",
      "min": 0,
      "max": 1
    }
  },
  "ray_resource": {
    "head_group_spec": {
      "spec_code": "string"
    },
    "worker_group_spec": [ {
      "name": "string",
      "spec_code": "string",
      "min_replicas": 10000,
      "max_replicas": 10000
    } ]
  },
  "caches": [ {
    "mnt_path": "string",
    "cache_ins_id": "string"
  } ],
  "error_code": "string",
  "error_msg": "string",
  "urls": [ {
    "name": "string",
    "address": "string",
    "type": "PRIVATE",
    "token": "string"
  } ],
  "business_engine_instance_ids": [ "string" ],
  "tokens_quota": {
    "total": 0,

```

```
"used" : 0,
"expire_time" : "2025-01-17T07:24:06.877Z"
},
"runtime_env_type" : "RELEASE",
"config" : {
  "ray_service_config" : {
    "log_config" : {
      "lts" : {
        "enabled" : false,
        "log_group_id" : "string",
        "log_stream_id" : "string"
      }
    },
    "ray_serve_config" : {
      "applications" : [ {
        "name" : "string",
        "import_path" : "string",
        "deployments" : [ {
          "name" : "string",
          "ray_actor_options" : {
            "num_cpus" : 0,
            "memory" : 0,
            "resources" : "string"
          },
          "autoscaling_config" : {
            "target_ongoing_requests" : 1000,
            "max_ongoing_requests" : 1000,
            "min_replicas" : 1000,
            "max_replicas" : 1000,
            "initial_replicas" : 1000,
            "upscale_delay_s" : 86400,
            "downscale_delay_s" : 86400
          },
          "num_replicas" : 1000,
          "user_config" : { },
          "max_replicas_per_node" : ""
        } ],
        "runtime_env" : {
          "working_dir" : "string",
          "env_vars" : {
            "additionalProp1" : "string",
            "additionalProp2" : "string",
            "additionalProp3" : "string"
          }
        }
      } ],
      "route_prefix" : "string"
    } ],
    "ray_cluster_config" : {
      "head_group_spec" : {
        "limits" : {
          "spec_code" : "string"
        },
        "requests" : {
          "spec_code" : "string"
        }
      },
      "worker_group_specs" : [ {
        "name" : "string",
        "min_replicas" : 10000,
        "max_replicas" : 10000,
        "limits" : {
          "spec_code" : "string"
        },
        "requests" : {
          "spec_code" : "string"
        }
      } ]
    },
  },
}
```

```
"data_infos" : [ {  
  "input_path" : "OBS://bucket/dir/",  
  "output_path" : "OBS://bucket/dir/",  
  "env_var_name" : "string"  
}]  
}  
}
```

状态码：400

BadRequest

```
{  
  "error_code" : "common.01000001",  
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,  
cause: TypeMismatchException"  
}
```

状态码：401

Unauthorized

```
{  
  "error_code" : "APIG.1002",  
  "error_msg" : "Incorrect token or token resolution failed"  
}
```

状态码：403

Forbidden

```
{  
  "error" : {  
    "code" : "403",  
    "message" : "X-Auth-Token is invalid in the request",  
    "title" : "Forbidden"  
  },  
  "error_code" : 403,  
  "error_msg" : "X-Auth-Token is invalid in the request",  
  "title" : "Forbidden"  
}
```

状态码：408

Request Time-out

```
{  
  "error_code" : "common.00000408",  
  "error_msg" : "timeout exception occurred"  
}
```

状态码：500

InternalServerError

```
{  
  "error_code" : "common.00000500",  
  "error_msg" : "internal error"  
}
```

状态码

状态码	描述
200	查询endpoint详情的响应。

状态码	描述
400	BadRequest
401	Unauthorized
403	Forbidden
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.6.4 修改端点

功能介绍

更新端点。用户可通过该接口更新某端点信息，输入为名称、描述、资源配置、镜像包配置、配置等信息，返回更新后的端点信息，rayservice的log_config会被忽略，无法更新。此接口为异步接口，配套使用接口ShowEndpoint查询端点信息，无特殊场景。

URI

PUT /v1/workspaces/{workspace_id}/endpoints/{endpoint_id}

表 4-167 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间ID，获取方法，请参见 获取工作空间ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
endpoint_id	是	String	参数解释： 端点ID。获取方法，请参见 获取端点ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-168 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户Token。用于调用接口的认证。获取方法，请参见认证鉴权。 约束限制：不涉及。 取值范围：长度不超过65534个字符。 默认取值：不涉及。

表 4-169 请求 Body 参数

参数	是否必选	参数类型	描述
name	否	String	参数解释：Endpoint名称。 约束限制：不涉及。 取值范围：长度为[1,128]包含中文、字母、数字、下划线、中划线、半角句号(.)、空格的组合。 默认取值：不涉及。
description	否	String	参数解释：描述信息。 约束限制：不涉及。 取值范围：[0,1024]。 默认取值：不涉及。
reserved_resource	否	ReservedResource object	参数解释：通用预留资源。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
ray_resource	否	RayResourceInput object	参数解释：创建Ray集群的资源配置。 约束限制：不涉及。
cap	否	CapRef object	参数解释：镜像包信息。 约束限制：不涉及。
config	否	EndpointConfig object	参数解释：端点的配置。 约束限制：不涉及。

参数	是否必选	参数类型	描述
public_access	否	Boolean	参数解释： 是否开启公网访问。 约束限制： 不涉及。 取值范围： 开启true，关闭false。 默认取值： 不涉及。

表 4-170 ReservedResource

参数	是否必选	参数类型	描述
apu	否	ResourceDef object	参数解释： 资源定义信息。 约束限制： 不涉及。
dpu	否	SqlResourceDef object	参数解释： Sql资源定义信息。 约束限制： 不涉及。
mu	否	ResourceDemand object	参数解释： 资源需求量配置。 约束限制： 不涉及。

表 4-171 ResourceDef

参数	是否必选	参数类型	描述
spec_code	是	String	参数解释： Endpoint状态。 约束限制： 资源规格编码，从查询规格列表ListSpecs接口获取。 取值范围： 不涉及。 默认取值： 不涉及。
warm_up_num	是	Integer	参数解释： 预热资源量。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
max_num	否	Integer	参数解释： 最大资源量，不填默认为预热资源量，即不使用弹性资源。 约束限制： 不涉及。 取值范围： 最小值为1，最大值为1000。 默认取值： 不涉及。

表 4-172 SqlResourceDef

参数	是否必选	参数类型	描述
spec_code	是	String	参数解释： Endpoint状态。 约束限制： 资源规格编码，从查询规格列表ListSpecs接口获取。 取值范围： 不涉及。 默认取值： 不涉及。
warm_up_num	是	Integer	参数解释： 预热资源量。 约束限制： 不涉及。 取值范围： 最小值为1，最大值为5000。 默认取值： 不涉及。
max_num	否	Integer	参数解释： 最大资源量，不填默认为预热资源量，即不使用弹性资源。 约束限制： 不涉及。 取值范围： 最小值为1，最大值为5000。 默认取值： 不涉及。
use_public_resource	否	Boolean	参数解释： 是否使用公共资源。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

表 4-173 RayResourceInput

参数	是否必选	参数类型	描述
head_node_resource	是	HeadNodeResourceDemand object	参数解释： headNode资源需求量配置。 约束限制： 不涉及。
work_node_resources	是	Array of WorkNodeResourceDemand objects	参数解释： 工作节点资源配置。 约束限制： 不涉及。 取值范围： [1,10]。 默认取值： 不涉及。

表 4-174 HeadNodeResourceDemand

参数	是否必选	参数类型	描述
spec_code	是	String	参数解释：资源规格。 约束限制：不涉及。 取值范围：从规格列表查询获取。 默认取值：不涉及。

表 4-175 WorkNodeResourceDemand

参数	是否必选	参数类型	描述
name	是	String	参数解释：资源配置名称。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
resource	是	ResourceDemand object	参数解释：资源需求量配置。 约束限制：不涉及。

表 4-176 ResourceDemand

参数	是否必选	参数类型	描述
spec_code	是	String	参数解释：资源规格。 约束限制：不涉及。 取值范围：从规格列表查询获取。 默认取值：不涉及。
min	是	Integer	参数解释：最小数。 约束限制：不涉及。 取值范围：最小值为0。 默认取值：不涉及。
max	是	Integer	参数解释：最大数。 约束限制：不涉及。 取值范围：最小值为1。 默认取值：不涉及。

表 4-177 CapRef

参数	是否必选	参数类型	描述
cap_id	否	String	参数解释：镜像包ID。 约束限制：不涉及。 取值范围：长度为[32,36]的英文、数字、中划线(-)的组合。 默认取值：不涉及。
version_id	否	String	参数解释：镜像包版本ID。 约束限制：不涉及。 取值范围：长度为[32,36]的英文、数字、中划线(-)的组合。 默认取值：不涉及。

表 4-178 EndpointConfig

参数	是否必选	参数类型	描述
ray_service_config	否	RayServiceConfig object	参数解释：RayService类型端点的配置。 约束限制：不涉及。

表 4-179 RayServiceConfig

参数	是否必选	参数类型	描述
ray_serve_config	是	RayServeConfig object	参数解释：RayServe配置。 约束限制：不涉及。
ray_cluster_config	是	RayClusterConfig object	参数解释：Ray集群配置。 约束限制：不涉及。
log_config	否	RayServiceLogConfig object	参数解释：RayService日志配置。 约束限制：不涉及。
data_infos	否	Array of DataBriefInfo objects	参数解释：数据信息。 约束限制：不涉及。 取值范围：[0,1]。 默认取值：不涉及。

表 4-180 RayServeConfig

参数	是否必选	参数类型	描述
applications	否	Array of ServeApplication objects	参数解释： Application。 约束限制： 不涉及。 取值范围： [0,5]。 默认取值： 不涉及。

表 4-181 ServeApplication

参数	是否必选	参数类型	描述
name	是	String	参数解释： Application的名称。 约束限制： 不涉及。 取值范围： 长度为[1,64]的中文、字母、数字、下划线、中划线、半角句号(.)、空格的组合。 默认取值： 不涉及。
import_path	是	String	参数解释： 输入路径。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
deployments	是	Array of Deployment objects	参数解释： Deployment列表。 约束限制： 不涉及。 取值范围： [1,5]。 默认取值： 不涉及。
runtime_env	是	ServeRuntimeEnv object	参数解释： Serve的运行时环境配置。 约束限制： 不涉及。 取值范围： 可选参数有： working_dir: 代码将在其中运行的工作目录。必须是远程URI，如s3或本地路径;env_vars: 要设置的环境变量。 默认取值： 不涉及。

参数	是否必选	参数类型	描述
route_prefix	否	String	参数解释： 路由前缀。 约束限制： 要求在RayServe中具有唯一性，不能跟其他的Application的前缀重复。 取值范围： 不涉及。 默认取值： 不涉及。

表 4-182 Deployment

参数	是否必选	参数类型	描述
name	否	String	参数解释： Deployment的名称。 约束限制： 不涉及。 取值范围： 长度为[1,64]的中文、字母、数字、下划线、中划线、半角句号(.)、空格的组合。 默认取值： 不涉及。
ray_actor_options	否	RayActorOptions object	参数解释： RayActor配置。 约束限制： 不涉及。
autoscaling_config	否	AutoscalingConfig object	参数解释： 自动扩缩的配置。 约束限制： 不涉及。
num_replicas	否	Integer	参数解释： 副本数量。 约束限制： 不涉及。 取值范围： [0,1000]。 默认取值： 不涉及。
user_config	否	Object	参数解释： 用户自定义配置。 约束限制： 不涉及。
max_replicas_per_node	否	Integer	参数解释： 每个节点允许的最大副本数。 约束限制： 不涉及。 取值范围： [1,100]。 默认取值： 1。
max_ongoing_requests	否	Integer	参数解释： 每个副本可接受的最大并发请求数。 约束限制： 不涉及。 取值范围： [1,100000]。 默认取值： 不涉及。

参数	是否必选	参数类型	描述
max_queued_requests	否	Integer	参数解释： deployment可接受的排队的最大请求数。 -1时表示无限制。 约束限制： 不涉及。 取值范围： [-1,100000]。 默认取值： 不涉及。

表 4-183 RayActorOptions

参数	是否必选	参数类型	描述
num_cpus	否	Double	参数解释： CPU数量。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
memory	否	Double	参数解释： 内存数量。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
resources	否	String	参数解释： 资源配置。 约束限制： 不涉及。 取值范围： 长度[0,1024]。 默认取值： 不涉及。

表 4-184 AutoscalingConfig

参数	是否必选	参数类型	描述
target_ongoing_requests	否	Integer	参数解释： 每个副本单位时间能提供的平均请求数。 约束限制： 不涉及。 取值范围： [1,1000]。 默认取值： 不涉及。
min_replicas	否	Integer	参数解释： 最小副本数。 约束限制： 不涉及。 取值范围： [0,1000]。 默认取值： 不涉及。

参数	是否必选	参数类型	描述
max_replicas	否	Integer	参数解释： 最大副本数。 约束限制： 不涉及。 取值范围： [0,1000]。 默认取值： 不涉及。
initial_replicas	否	Integer	参数解释： 初始副本数。 约束限制： 不涉及。 取值范围： [0,1000]。 默认取值： 不涉及。
upscale_delay_s	否	Integer	参数解释： 扩容之前的等待时间。 约束限制： 不涉及。 取值范围： [0,86400]。 默认取值： 不涉及。
downscale_delay_s	否	Integer	参数解释： 缩容之前的等待时间。 约束限制： 不涉及。 取值范围： [0,86400]。 默认取值： 不涉及。

表 4-185 ServeRuntimeEnv

参数	是否必选	参数类型	描述
working_dir	否	String	参数解释： 代码将在其中运行的工作目录。 约束限制： 必须是远程URI，如s3或本地路径。 取值范围： 不涉及。 默认取值： 不涉及。
env_vars	否	Map<String,String>	参数解释： 要设置的环境变量。 约束限制： 不涉及。

表 4-186 RayClusterConfig

参数	是否必选	参数类型	描述
enable_custom	否	Boolean	参数解释：是否开启资源自定义。 约束限制：不涉及。 取值范围：开启true，关闭false。 默认取值：false。
head_group_spec	是	HeadGroupSpecV2 object	参数解释：HeadGroup的配置。 约束限制：不涉及。
worker_group_specs	是	Array of WorkerGroupSpecV2 objects	参数解释：Worker Group的配置。 约束限制：[1,1000]。

表 4-187 HeadGroupSpecV2

参数	是否必选	参数类型	描述
limits	否	ResourceSpec object	参数解释：资源配置。 约束限制：不涉及。
requests	否	ResourceSpec object	参数解释：资源配置。 约束限制：不涉及。

表 4-188 WorkerGroupSpecV2

参数	是否必选	参数类型	描述
name	否	String	参数解释：名称。 约束限制：不涉及。 取值范围：长度[0,128]字母、数字、下划线(_)、中划线(-)的组合。 默认取值：不涉及。
min_replicas	否	Integer	参数解释：最小副本数。 约束限制：不涉及。 取值范围：[0,10000]。 默认取值：不涉及。

参数	是否必选	参数类型	描述
max_replicas	否	Integer	参数解释：最大副本数。 约束限制：不涉及。 取值范围：[1,10000]。 默认取值：不涉及。
limits	否	ResourceSpec object	参数解释：资源限制。 约束限制：不涉及。
requests	否	ResourceSpec object	参数解释：最小资源数。 约束限制：不涉及。

表 4-189 ResourceSpec

参数	是否必选	参数类型	描述
spec_code	否	String	参数解释：资源规格。 约束限制：不涉及。 取值范围：从规格列表查询获取。 默认取值：不涉及。
cpu_resource	否	Integer	参数解释：cpu核数量。 约束限制：不涉及。 取值范围：[0,192]。 默认取值：null。
memory_resource	否	Integer	参数解释：内存大小。 约束限制：不涉及。 取值范围：[0,1536]。 默认取值：null。
npu_resource	否	Integer	参数解释：昇腾卡数量。 约束限制：不涉及。 取值范围：[0,8]。 默认取值：null。

表 4-190 RayServiceLogConfig

参数	是否必选	参数类型	描述
lts	否	RayServiceLogLtsConfig object	参数解释：RayService日志配置。 约束限制：不涉及。

表 4-191 RayServiceLogLtsConfig

参数	是否必选	参数类型	描述
enabled	否	Boolean	参数解释：是否开启RayService日志转储到LTS。 约束限制：不涉及。 取值范围：true，false。 默认取值：false。
log_group_id	否	String	参数解释：RayService日志转储到的LTS日志组id。 约束限制：需要是已经存在的LTS日志组id。 取值范围：不涉及。 默认取值：不涉及。
log_stream_id	否	String	参数解释：RayService日志转储到的LTS日志流id。 约束限制：需要是已经存在的LTS日志流id。 取值范围：不涉及。 默认取值：不涉及

表 4-192 DataBriefInfo

参数	是否必选	参数类型	描述
input_path	否	String	参数解释：地址。 约束限制：不涉及。 取值范围：长度[4,1024]。 默认取值：不涉及。

响应参数

状态码：202
查询endpoint详情的响应。

状态码：400

表 4-193 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-194 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-195 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-196 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

更新端点。具体更新信息如下样例所示。

```
PUT https://{endpoint}/v1/workspaces/{workspace_id}/endpoints/{endpoint_id}
{
```

```
"name" : "endpoint1",
"description" : "description",
"reserved_resource" : {
  "apu" : {
    "spec_code" : "dpu.d1x",
    "warm_up_num" : 1,
    "max_num" : 100
  },
  "dpu" : {
    "spec_code" : "dpu.d1x",
    "warm_up_num" : 1,
    "max_num" : 100,
    "use_public_resource" : false
  },
  "mu" : {
    "spec_code" : "string",
    "min" : 0,
    "max" : 1
  }
},
"ray_resource" : {
  "head_node_resource" : {
    "spec_code" : "string"
  },
  "work_node_resources" : [ {
    "name" : "cpuGroup1",
    "resource" : {
      "spec_code" : "string",
      "min" : 0,
      "max" : 1
    }
  }
],
"cap" : {
  "cap_id" : "0b5633ba2b904511ad514346f4d23d4b",
  "version_id" : "0b5633ba2b904511ad514346f4d23d4b"
},
"config" : {
  "ray_service_config" : {
    "ray_serve_config" : {
      "applications" : [ {
        "name" : "string",
        "import_path" : "string",
        "deployments" : [ {
          "name" : "string",
          "ray_actor_options" : {
            "num_cpus" : 0,
            "memory" : 0,
            "resources" : "string"
          }
        }
      ],
      "autoscaling_config" : {
        "target_ongoing_requests" : 1000,
        "max_ongoing_requests" : 1000,
        "min_replicas" : 1000,
        "max_replicas" : 1000,
        "initial_replicas" : 1000,
        "upscale_delay_s" : 86400,
        "downscale_delay_s" : 86400
      },
      "num_replicas" : 1000,
      "user_config" : { },
      "max_replicas_per_node" : 1
    }
  ],
  "runtime_env" : {
    "working_dir" : "string",
    "env_vars" : {
      "additionalProp1" : "string",
      "additionalProp2" : "string",
      "additionalProp3" : "string"
```

```
    }
  },
  "route_prefix" : "string"
} ]
},
"ray_cluster_config" : {
  "head_group_spec" : {
    "limits" : {
      "spec_code" : "string"
    },
    "requests" : {
      "spec_code" : "string"
    }
  },
  "worker_group_specs" : [ {
    "name" : "string",
    "min_replicas" : 10000,
    "max_replicas" : 10000,
    "limits" : {
      "spec_code" : "string"
    },
    "requests" : {
      "spec_code" : "string"
    }
  } ]
},
"data_infos" : [ {
  "input_path" : "OBS://bucket/dir/"
} ]
}
}
```

响应示例

状态码：202

查询endpoint详情的响应。

```
{
  "visibility" : "PRIVATE",
  "id" : "0b5633ba2b904511ad514346f4d23d4b",
  "name" : "endpoint1",
  "type" : "job",
  "status" : "CREATING",
  "description" : "description",
  "create_time" : "2023-05-30T12:24:30.401Z",
  "update_time" : "2023-05-30T12:24:30.401Z",
  "owner" : {
    "domain_name" : "string",
    "domain_id" : "0b5633ba2b904511ad514346f4d23d4b",
    "user_name" : "string",
    "user_id" : "0b5633ba2b904511ad514346f4d23d4b"
  },
  "cap" : {
    "cap_id" : "0b5633ba2b904511ad514346f4d23d4b",
    "version_id" : "0b5633ba2b904511ad514346f4d23d4b"
  },
  "reserved_resource" : {
    "apu" : {
      "spec_code" : "dpu.d1x",
      "warm_up_num" : 1,
      "max_num" : 100
    },
    "dpu" : {
      "spec_code" : "dpu.d1x",
      "warm_up_num" : 1,
      "max_num" : 100,
      "use_public_resource" : false
    }
  }
}
```

```
    },
    "mu" : {
      "spec_code" : "string",
      "min" : 0,
      "max" : 1
    }
  },
  "ray_resource" : {
    "head_group_spec" : {
      "spec_code" : "string"
    },
    "worker_group_spec" : [ {
      "name" : "string",
      "spec_code" : "string",
      "min_replicas" : 10000,
      "max_replicas" : 10000
    } ]
  },
  "caches" : [ {
    "mnt_path" : "string",
    "cache_ins_id" : "string"
  } ],
  "error_code" : "string",
  "error_msg" : "string",
  "urls" : [ {
    "name" : "string",
    "address" : "string",
    "type" : "PRIVATE",
    "token" : "string"
  } ],
  "business_engine_instance_ids" : [ "string" ],
  "tokens_quota" : {
    "total" : 0,
    "used" : 0,
    "expire_time" : "2025-01-17T07:30:35.677Z"
  },
  "runtime_env_type" : "RELEASE",
  "config" : {
    "ray_service_config" : {
      "ray_serve_config" : {
        "applications" : [ {
          "name" : "string",
          "import_path" : "string",
          "deployments" : [ {
            "name" : "string",
            "ray_actor_options" : {
              "num_cpus" : 0,
              "memory" : 0,
              "resources" : "string"
            }
          } ],
        } ],
      },
      "autoscaling_config" : {
        "target_ongoing_requests" : 1000,
        "max_ongoing_requests" : 1000,
        "min_replicas" : 1000,
        "max_replicas" : 1000,
        "initial_replicas" : 1000,
        "upscale_delay_s" : 86400,
        "downscale_delay_s" : 86400
      },
      "num_replicas" : 1000,
      "user_config" : { },
      "max_replicas_per_node" : ""
    } ],
    "runtime_env" : {
      "working_dir" : "string",
      "env_vars" : {
        "additionalProp1" : "string",
        "additionalProp2" : "string",
        "additionalProp3" : "string"
      }
    }
  }
}
```

```
    }
  },
  "route_prefix" : "string"
}]
},
"ray_cluster_config" : {
  "head_group_spec" : {
    "limits" : {
      "spec_code" : "string"
    },
    "requests" : {
      "spec_code" : "string"
    }
  },
  "worker_group_specs" : [ {
    "name" : "string",
    "min_replicas" : 10000,
    "max_replicas" : 10000,
    "limits" : {
      "spec_code" : "string"
    },
    "requests" : {
      "spec_code" : "string"
    }
  }
],
"data_infos" : [ {
  "input_path" : "OBS://bucket/dir/",
  "output_path" : "OBS://bucket/dir/",
  "env_var_name" : "string"
}
]
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch., cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
202	查询endpoint详情的响应。
400	BadRequest
401	Unauthorized
403	Forbidden
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.6.5 删除端点

功能介绍

删除端点。用户可通过该接口删除工作空间下的端点，输入端点id，返回操作结果。此接口为异步接口，配套使用接口ShowEndpoint查询端点信息，无特殊场景。

URI

DELETE /v1/workspaces/{workspace_id}/endpoints/{endpoint_id}

表 4-197 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间ID，获取方法，请参见 获取工作空间ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
endpoint_id	是	String	参数解释： 端点ID。获取方法，请参见 获取端点ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

表 4-198 Query 参数

参数	是否必选	参数类型	描述
is_delete_app_instance	否	Boolean	参数解释： 删除ray endpoint时是否删除ray endpoint所有job运行实例。 约束限制： 不涉及。 取值范围： true, false。 默认取值： false。

请求参数

表 4-199 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

响应参数

状态码：202

表 4-200 响应 Body 参数

参数	参数类型	描述
id	String	参数解释：端点ID。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

状态码：400

表 4-201 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-202 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408**表 4-203** 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500**表 4-204** 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

```
DELETE https://{endpoint}/v1/workspaces/{workspace_id}/endpoints/{endpoint_id}
```

响应示例

状态码：202

删除Endpoint的响应。

```
{
  "id" : "ac8111bf-3601-4905-8ddd-b41d3e636a4e"
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch., cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
202	删除Endpoint的响应。
400	BadRequest
401	Unauthorized
403	Forbidden
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.6.6 订阅端点

功能介绍

开通端点。用户可通过此接口开通公共端点，输入为端点id，返回操作结果。此接口为同步接口，无配套使用接口和特殊场景。

URI

POST /v1/workspaces/{workspace_id}/endpoints/{endpoint_id}/subscribe

表 4-205 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： workspaceID，获取方法，请参见 获取workspaceID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
endpoint_id	是	String	参数解释： 端点ID。获取方法，请参见 获取端点ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-206 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

响应参数

状态码：200

表 4-207 响应 Body 参数

参数	参数类型	描述
-	AnyType	OK

状态码：400

表 4-208 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-209 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-210 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-211 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-212 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

PUT https://{endpoint}/v1/workspaces/{workspace_id}/endpoints/{endpoint_id}/subscribe

响应示例

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	OK
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.7 模型定义

4.7.1 创建模型

功能介绍

创建模型。用户可通过该接口创建模型，支持传入名称、描述、版本名称、版本描述、基模型类型和模型所在的OBS路径，返回模型ID及其版本ID。此接口为异步接

口，配套使用接口ListModels查询模型列表。存在同名校验，不支持创建重复名称的模型。

URI

POST /v1/workspaces/{workspace_id}/models

表 4-213 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释：工作空间ID。获取方法，请参见获取工作空间ID。 约束限制：不涉及。 取值范围：长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值：不涉及。

请求参数

表 4-214 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户Token。用于调用接口的认证。获取方法，请参见认证鉴权。 约束限制：不涉及。 取值范围：长度不超过65534个字符。 默认取值：不涉及。

表 4-215 请求 Body 参数

参数	是否必选	参数类型	描述
name	是	String	参数解释：一个Model的名称。 约束限制：不涉及。 取值范围：长度为[1,64]的中文、字母、数字、下划线(_)、中划线(-)、半角句号(.)、空格的组合。 默认取值：不涉及。

参数	是否必选	参数类型	描述
description	否	String	参数解释：描述信息。 约束限制：不涉及。 取值范围：[0,1024]。 默认取值：不涉及。
type	是	String	参数解释：模型的类型。 约束限制：不涉及。 取值范围：LLM_MODEL：大语言模型。 默认取值：不涉及。
version	是	ModelVersionInput object	参数解释：模型每个版本的详细信息。 约束限制：不涉及。

表 4-216 ModelVersionInput

参数	是否必选	参数类型	描述
name	是	String	参数解释：模型版本名称。 约束限制：不涉及。 取值范围：长度为[1,64]的中文、字母、数字、下划线(_)、中划线(-)、半角句号(.)、空格的组合。 默认取值：不涉及。
description	否	String	参数解释：描述信息。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
config	否	ModelConfig object	参数解释：模型配置。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

表 4-217 ModelConfig

参数	是否必选	参数类型	描述
llm_model_config	否	LlmModelConfig object	参数解释：Llm模型配置。 约束限制：不涉及。

表 4-218 LlmModelConfig

参数	是否必选	参数类型	描述
base_model_type	是	String	参数解释： 基模型类型。 约束限制： 不涉及。 取值范围： 请从 列举基模型 接口响应中获取。 默认取值： 不涉及。
model_path	否	String	参数解释： 模型文件路径。 约束限制： 合法OBS路径。 取值范围： 不涉及。 默认取值： 不涉及。
local_model_path	否	String	参数解释： 本地模型文件路径。 约束限制： 本地模型文件路径。 取值范围： 不涉及。 默认取值： 不涉及。

响应参数

状态码：200

表 4-219 响应 Body 参数

参数	参数类型	描述
id	String	参数解释： 模型ID。获取方法，请参见 获取模型ID 。 约束限制： 不涉及。 取值范围： 长度为[32,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
version_id	String	参数解释： 模型版本ID。 约束限制： 系统自动生成无法修改，输入不生效。 取值范围： 长度为[32,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

状态码：400

表 4-220 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-221 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-222 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-223 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-224 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

创建模型接口。名称为LLama3-8b，描述为description，类型为LLM_MODEL，版本具体信息如下请求样例所示。

```
POST https://{endpoint}/v1/workspaces/{workspace_id}/models
{
  "name": "LLama3-8b",
  "description": "description",
  "type": "LLM_MODEL",
  "version": {
    "name": "name",
    "description": "description",
    "config": {
      "llm_model_config": {
        "base_model_type": "string",
        "model_path": "obs://bucketname/objectdir/modelpath/"
      }
    }
  }
}
```

响应示例

状态码：200

创建Model的响应体。

```
{
  "id": "ac8111bf-3601-4905-8ddd-b41d3e636a4e"
}
```

状态码：400

BadRequest

```
{
  "error_code": "common.01000001",
}
```

```
"error_msg": "failed to read http request, please check your input, code: 400, reason: Type mismatch.,  
cause: TypeMismatchException"  
}
```

状态码：401

Unauthorized

```
{  
  "error_code": "APIG.1002",  
  "error_msg": "Incorrect token or token resolution failed"  
}
```

状态码：403

Forbidden

```
{  
  "error": {  
    "code": "403",  
    "message": "X-Auth-Token is invalid in the request",  
    "title": "Forbidden"  
  },  
  "error_code": 403,  
  "error_msg": "X-Auth-Token is invalid in the request",  
  "title": "Forbidden"  
}
```

状态码：404

NotFound

```
{  
  "error_code": "common.01000001",  
  "error_msg": "response status exception, code: 404"  
}
```

状态码：408

Request Time-out

```
{  
  "error_code": "common.00000408",  
  "error_msg": "timeout exception occurred"  
}
```

状态码：500

InternalServerError

```
{  
  "error_code": "common.00000500",  
  "error_msg": "internal error"  
}
```

状态码

状态码	描述
200	创建Model的响应体。
400	BadRequest
401	Unauthorized
403	Forbidden

状态码	描述
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.7.2 查询模型列表

功能介绍

列举模型。用户可通过该接口查询模型列表，接口支持分页查询、支持通过名称、ID和类型查询。支持排序。返回为符合过滤条件的模型列表。此接口为同步接口，无配套使用接口和特殊场景。

URI

GET /v1/workspaces/{workspace_id}/models

表 4-225 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间ID。获取方法，请参见 获取工作空间ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

表 4-226 Query 参数

参数	是否必选	参数类型	描述
offset	否	Integer	参数解释： 偏移量，表示从此偏移量开始查询。 约束限制： 不涉及。 取值范围： [0,1000000000]。 默认取值： 0。

参数	是否必选	参数类型	描述
limit	否	Integer	参数解释： 指定每一页返回的最大条目数。 约束限制： 不涉及。 取值范围： [1,100]。 默认取值： 10。
name	否	String	参数解释： 通过名字搜索 Model，支持模糊查询。 约束限制： 不涉及。 取值范围： 长度为[1,128]的英文字符、数字、中划线(-)和下划线(_)的组合 默认取值： 不涉及。
id	否	String	参数解释： 通过模型ID检索。 约束限制： 不涉及。 取值范围： 长度为[32,36]的英文字符、数字和中划线(-)的组合，请参见 获取模型ID 。 默认取值： 不涉及。
type	否	String	参数解释： 通过模型类型检索。 约束限制： 不涉及。 取值范围： 长度为[1,128]个字符。 默认取值： 不涉及。
visibility	否	String	参数解释： 可见性检索的参数。 约束限制： 不涉及。 取值范围： • PRIVATE：私有，用户自己创建的； • PUBLIC：公共，查询所有公共的，包括其他用户创建的； • ALL：所有的。 默认取值： 默认为空，不填表示不限制，则查出当前用户下的，包括PRIVATE和PUBLIC，不包括其他用户创建的。

参数	是否必选	参数类型	描述
sort_by	否	String	参数解释： 指定排序字段。 约束限制： 不涉及。 取值范围： • CREATE_TIME：创建时间，默认值； • UPDATE_TIME：更新时间； • NAME：服务名称。 默认取值： 不涉及。
order_by	否	String	参数解释： 排序方式。 约束限制： 不涉及。 取值范围： • ASC：递增排序； • DESC：递减排序。 默认取值： DESC。

请求参数

表 4-227 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

响应参数

状态码：200

表 4-228 响应 Body 参数

参数	参数类型	描述
total	Integer	参数解释：符合条件的总数。 约束限制：不涉及。 取值范围：[0,10000]。 默认取值：不涉及。
models	Array of ModelInfo objects	参数解释：符合条件的总数。 约束限制：[0,1000]。

表 4-229 ModelInfo

参数	参数类型	描述
visibility	String	参数解释：可见性。 约束限制：不涉及。 取值范围： - PRIVATE：私有 - PUBLIC：公共 默认取值：PRIVATE。
id	String	参数解释：模型ID。获取方法，请参见 获取模型ID 。 约束限制：不涉及。 取值范围：长度为[32,36]的英文字符、数字和中划线(-)的组合。 默认取值：不涉及。
name	String	参数解释：一个Model的名称。 约束限制：不涉及。 取值范围：长度为[1,64]的中文、字母、数字、下划线(_)、中划线(-)、半角句号(.)、空格的组合。 默认取值：不涉及。
description	String	参数解释：描述信息。 约束限制：不涉及。 取值范围：[0,1024]。 默认取值：不涉及。

参数	参数类型	描述
type	String	参数解释：模型的类型。 约束限制：不涉及。 取值范围：LLM_MODEL：大语言模型。 默认取值：不涉及。
create_time	String	参数解释：创建时间。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
update_time	String	参数解释：更新时间。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
current_version	ModelVersionInfo object	参数解释：模型版本详情信息。 约束限制：不涉及。
create_user	User object	参数解释：租户信息内容，包含账户名称、账户ID和用户名称、用户ID。 约束限制：不涉及。
update_user	User object	参数解释：租户信息内容，包含账户名称、账户ID和用户名称、用户ID。 约束限制：不涉及。

表 4-230 ModelVersionInfo

参数	参数类型	描述
id	String	参数解释：模型版本ID。 约束限制：系统自动生成无法修改，输入不生效。 取值范围：长度为[32,36]的英文字符、数字和中划线(-)的组合。 默认取值：不涉及。
name	String	参数解释：模型版本名称。 约束限制：不涉及。 取值范围：长度为[1,64]的中文、字母、数字、下划线(_)、中划线(-)、半角句号(.)、空格的组合。 默认取值：不涉及。

参数	参数类型	描述
cap_white_list	Array of strings	参数解释：镜像包白名单。 约束限制：[0,100]。
description	String	参数解释：描述信息。 约束限制：不涉及。 取值范围：[0,1024]。 默认取值：不涉及。
config	ModelConfig object	参数解释：模型配置。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
create_time	String	参数解释：创建时间。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
create_user	User object	参数解释：租户信息内容，包含账户名称、账户ID和用户名称、用户ID。 约束限制：不涉及。

表 4-231 ModelConfig

参数	参数类型	描述
llm_model_config	LlmModelConfig object	参数解释：Llm模型配置。 约束限制：不涉及。

表 4-232 LlmModelConfig

参数	参数类型	描述
base_model_type	String	参数解释：基模型类型。 约束限制：不涉及。 取值范围：请从 列举基模型 接口响应中获取。 默认取值：不涉及。
model_path	String	参数解释：模型文件路径。 约束限制：合法OBS路径。 取值范围：不涉及。 默认取值：不涉及。

参数	参数类型	描述
local_model_path	String	参数解释：本地模型文件路径。 约束限制：本地模型文件路径。 取值范围：不涉及。 默认取值：不涉及。

表 4-233 User

参数	参数类型	描述
domain_name	String	参数解释：账户名称。 约束限制：不涉及。 取值范围：长度为[0,64]个字符。 默认取值：不涉及。
domain_id	String	参数解释：账号ID。 约束限制：不涉及。 取值范围：长度为[1,36]的数字、英文字符和中划线(-)的组合。 默认取值：不涉及。
user_name	String	参数解释：用户名称。 约束限制：不涉及。 取值范围：长度为[0,64]个字符。 默认取值：不涉及。
user_id	String	参数解释：用户ID。 约束限制：不涉及。 取值范围：长度为[1,36]的数字、英文字符和中划线(-)的组合。 默认取值：不涉及。

状态码：400

表 4-234 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401**表 4-235** 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404**表 4-236** 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-237 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-238 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释： 错误描述。 约束限制： 不涉及。 取值范围： 长度为[2,512]个字符。 默认取值： 不涉及。
solution_msg	String	参数解释： 解决方案描述。 约束限制： 不涉及。 取值范围： 长度为[0,4096]个字符。 默认取值： 不涉及。

请求示例

GET https://{endpoint}/v1/workspaces/{workspace_id}/models

响应示例

状态码：200
查询模型列表的响应体。

```
{
  "total": 10,
  "models": [ {
    "visibility": "PRIVATE",
    "id": "0b5633ba2b904511ad514346f4d23d4b",
    "name": "LLama3-8b",
    "description": "description",
    "type": "LLM_MODEL",
    "create_time": "2023-05-30T12:24:30.401Z",
    "update_time": "2023-05-30T12:24:30.401Z",
    "current_version": {
      "id": "0b5633ba2b904511ad514346f4d23d4b",
      "name": "name",
      "cap_white_list": [ "cap1 cap2 ca* cap*" ],
      "description": "description",
      "config": {
        "llm_model_config": {
          "base_model_type": "string",
          "model_path": "obs://bucketname/objectdir/modelpath/"
        }
      },
      "create_time": "2023-05-30T12:24:30.401Z",
      "create_user": {
        "domain_name": "string",
        "domain_id": "0b5633ba2b904511ad514346f4d23d4b",
        "user_name": "string",
        "user_id": "0b5633ba2b904511ad514346f4d23d4b"
      }
    },
    "create_user": {
      "domain_name": "string",
      "domain_id": "0b5633ba2b904511ad514346f4d23d4b",
      "user_name": "string",
      "user_id": "0b5633ba2b904511ad514346f4d23d4b"
    },
    "update_user": {
      "domain_name": "string",
      "domain_id": "0b5633ba2b904511ad514346f4d23d4b",

```

```
{
  "user_name" : "string",
  "user_id" : "0b5633ba2b904511ad514346f4d23d4b"
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	查询模型列表的响应体。
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.7.3 列举基模型

功能介绍

列举基模型。用户可通过该接口查询基模型列表，接口支持分页查询。返回为符合条件的基模型列表。此接口为同步接口，无配套使用接口和特殊场景。

URI

GET /v1/base-models

表 4-239 Query 参数

参数	是否必选	参数类型	描述
offset	否	Integer	参数解释： 偏移量，表示从此偏移量开始查询。 约束限制： 不涉及。 取值范围： [0,10000000000]。 默认取值： 0。
limit	否	Integer	参数解释： 指定每一页返回的最大条目数。 约束限制： 不涉及。 取值范围： [1,100]。 默认取值： 10。

请求参数

表 4-240 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户Token。用于调用接口的认证。获取方法，请参见认证鉴权。 约束限制：不涉及。 取值范围：长度不超过65534个字符。 默认取值：不涉及。

响应参数

状态码：200

表 4-241 响应 Body 参数

参数	参数类型	描述
total	Integer	参数解释：符合条件的总数。 约束限制：不涉及。 取值范围：[0,10000]。 默认取值：不涉及。
models	Array of BaseModel objects	参数解释：列表信息。 约束限制：[0,1000]。

表 4-242 BaseModel

参数	参数类型	描述
base_model_type	String	参数解释：基模型类型。 约束限制：不涉及。 取值范围：请从列举基模型接口响应中获取。 默认取值：不涉及。
max_tokens	Integer	参数解释：支持的最大Token数。 约束限制：不涉及。 取值范围：最小值1。 默认取值：不涉及。

参数	参数类型	描述
num_of_mu	Integer	参数解释：消耗MU数量。 约束限制：不涉及。 取值范围：[1,128]。 默认取值：不涉及。

状态码：400

表 4-243 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-244 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-245 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-246 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-247 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

GET https://{endpoint}/v1/base-models

响应示例

状态码：200

查询基模型列表的响应体。

```
{
  "total" : 1,
  "models" : [ {
    "base_model_type" : "string",
    "max_tokens" : 1,
    "num_of_mu" : 128
  } ]
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error": {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	查询基模型列表的响应体。
400	BadRequest
401	Unauthorized

状态码	描述
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.7.4 删除模型的指定版本

功能介绍

删除模型的指定版本。如果该版本正在被使用，则无法删除，需要删除使用该模型的推理服务才能删除。用户可通过该接口删除模型版本。此接口为同步接口，无配套使用接口。

URI

DELETE /v1/workspaces/{workspace_id}/models/{model_id}/versions/{version_id}

表 4-248 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间ID。获取方法，请参见 获取工作空间ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
model_id	是	String	参数解释： 模型ID。获取方法，请参见 获取模型ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

参数	是否必选	参数类型	描述
version_id	是	String	参数解释： 版本ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-249 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

响应参数

状态码：202

表 4-250 响应 Body 参数

参数	参数类型	描述
-	AnyType	OK

状态码：400

表 4-251 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释： 错误码。 约束限制： 不涉及。 取值范围： 长度为[8,36]个字符。 默认取值： 不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-252 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-253 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408**表 4-254 响应 Body 参数**

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500**表 4-255 响应 Body 参数**

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

DELETE https://{endpoint}/v1/workspaces/{workspace_id}/models/{model_id}/versions/{version_id}

响应示例

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error": {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
202	OK
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.7.5 更新模型信息

功能介绍

更新模型，会生成新版本。通过该接口可以更新模型信息，包括名称、描述、当前版本ID。支持新增版本，支持传入版本名称、描述、基模型类型和模型地址。此接口为异步接口，配套使用接口ListModelVersions查看模型版本信息。如果版本是失败的，则不允许更新为当前版本。新增模型时模型路径需要存在，否则会报错。

URI

PUT /v1/workspaces/{workspace_id}/models/{model_id}

表 4-256 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间ID。获取方法，请参见 获取工作空间ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
model_id	是	String	参数解释： 模型ID。获取方法，请参见 获取模型ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-257 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

表 4-258 请求 Body 参数

参数	是否必选	参数类型	描述
name	否	String	参数解释： 一个Model的名称。 约束限制： 不涉及。 取值范围： 长度为[1,64]的中文、字母、数字、下划线(_)、中划线(-)、半角句号(.)、空格的组合。 默认取值： 不涉及。

参数	是否必选	参数类型	描述
description	否	String	参数解释： 描述信息。 约束限制： 不涉及。 取值范围： [0,1024]。 默认取值： 不涉及。
current_version_id	否	String	参数解释： 模型版本ID。 约束限制： 系统自动生成无法修改，输入不生效。 取值范围： 长度为[32,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
version	否	ModelVersionInput object	参数解释： 模型每个版本的详细信息。 约束限制： 不涉及。

表 4-259 ModelVersionInput

参数	是否必选	参数类型	描述
name	是	String	参数解释： 模型版本名称。 约束限制： 不涉及。 取值范围： 长度为[1,64]的中文、字母、数字、下划线(_)、中划线(-)、半角句号(.)、空格的组合。 默认取值： 不涉及。
description	否	String	参数解释： 描述信息。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
config	否	ModelConfig object	参数解释： 模型配置。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

表 4-260 ModelConfig

参数	是否必选	参数类型	描述
llm_model_config	否	LlmModelConfig object	参数解释：Llm模型配置。 约束限制：不涉及。

表 4-261 LlmModelConfig

参数	是否必选	参数类型	描述
base_model_type	是	String	参数解释：基模型类型。 约束限制：不涉及。 取值范围：请从 列举基模型 接口响应中获取。 默认取值：不涉及。
model_path	否	String	参数解释：模型文件路径。 约束限制：合法OBS路径。 取值范围：不涉及。 默认取值：不涉及。
local_model_path	否	String	参数解释：本地模型文件路径。 约束限制：本地模型文件路径。 取值范围：不涉及。 默认取值：不涉及。

响应参数

状态码：200

OK。

状态码：400

表 4-262 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-263 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-264 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408**表 4-265 响应 Body 参数**

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500**表 4-266 响应 Body 参数**

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释： 错误描述。 约束限制： 不涉及。 取值范围： 长度为[2,512]个字符。 默认取值： 不涉及。
solution_msg	String	参数解释： 解决方案描述。 约束限制： 不涉及。 取值范围： 长度为[0,4096]个字符。 默认取值： 不涉及。

请求示例

更新模型。名称为LLama3-8b，描述为description，当前版本ID为0b5633ba2b904511ad514346f4d23d4b，生成新版本配置如下样例所示。

```
PUT https://{endpoint}/v1/workspaces/{workspace_id}/models/{model_id}

{
  "name": "LLama3-8b",
  "description": "description",
  "current_version_id": "0b5633ba2b904511ad514346f4d23d4b",
  "version": {
    "name": "name",
    "description": "description",
    "config": {
      "llm_model_config": {
        "base_model_type": "string",
        "model_path": "obs://bucketname/objectdir/modelpath/"
      }
    }
  }
}
```

响应示例

状态码：200

OK。

```
{
  "visibility": "PRIVATE",
  "id": "0b5633ba2b904511ad514346f4d23d4b",
  "name": "LLama3-8b",
  "description": "description",
  "type": "LLM_MODEL",
  "create_time": "2023-05-30T12:24:30.401Z",
  "update_time": "2023-05-30T12:24:30.401Z",
  "current_version": {
    "id": "0b5633ba2b904511ad514346f4d23d4b",
    "name": "name",
    "cap_white_list": [ "cap1 cap2 ca* cap*" ],
    "description": "description",
    "config": {
      "llm_model_config": {
        "base_model_type": "string",
        "model_path": "obs://bucketname/objectdir/modelpath/"
      }
    }
  }
}
```

```
    },
    "create_time": "2023-05-30T12:24:30.401Z",
    "create_user": {
      "domain_name": "string",
      "domain_id": "0b5633ba2b904511ad514346f4d23d4b",
      "user_name": "string",
      "user_id": "0b5633ba2b904511ad514346f4d23d4b"
    }
  },
  "create_user": {
    "domain_name": "string",
    "domain_id": "0b5633ba2b904511ad514346f4d23d4b",
    "user_name": "string",
    "user_id": "0b5633ba2b904511ad514346f4d23d4b"
  },
  "update_user": {
    "domain_name": "string",
    "domain_id": "0b5633ba2b904511ad514346f4d23d4b",
    "user_name": "string",
    "user_id": "0b5633ba2b904511ad514346f4d23d4b"
  },
  "new_version": {
    "id": "0b5633ba2b904511ad514346f4d23d4b",
    "name": "name",
    "cap_white_list": [ "cap1 cap2 ca* cap*" ],
    "description": "description",
    "config": {
      "llm_model_config": {
        "base_model_type": "string",
        "model_path": "obs://bucketname/objectdir/modelpath/"
      }
    }
  },
  "create_time": "2023-05-30T12:24:30.401Z",
  "create_user": {
    "domain_name": "string",
    "domain_id": "0b5633ba2b904511ad514346f4d23d4b",
    "user_name": "string",
    "user_id": "0b5633ba2b904511ad514346f4d23d4b"
  }
}
```

状态码：400

BadRequest

```
{
  "error_code": "common.01000001",
  "error_msg": "failed to read http request, please check your input, code: 400, reason: Type mismatch, cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code": "APIG.1002",
  "error_msg": "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error": {
    "code": "403",
    "message": "X-Auth-Token is invalid in the request",
    "title": "Forbidden"
  }
}
```

```
},
"error_code" : 403,
"error_msg" : "X-Auth-Token is invalid in the request",
"title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
"error_code" : "common.01000001",
"error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
"error_code" : "common.00000408",
"error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
"error_code" : "common.00000500",
"error_msg" : "internal error"
}
```

状态码

状态码	描述
200	OK。
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.7.6 查看模型版本列表

功能介绍

查询模型的版本列表。用户可通过该接口查询模型版本列表，接口支持分页查询、支持通过名称、ID查询。返回为符合过滤条件的模型版本列表。此接口为同步接口，无配套使用接口和特殊场景。

URI

GET /v1/workspaces/{workspace_id}/models/{model_id}/versions

表 4-267 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间ID。获取方法，请参见 获取工作空间ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
model_id	是	String	参数解释： 模型ID。获取方法，请参见 获取模型ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

表 4-268 Query 参数

参数	是否必选	参数类型	描述
offset	否	Integer	参数解释： 偏移量，表示从此偏移量开始查询。 约束限制： 不涉及。 取值范围： [0,1000000000]。 默认取值： 0。
limit	否	Integer	参数解释： 指定每一页返回的最大条目数。 约束限制： 不涉及。 取值范围： [1,100]。 默认取值： 10。

参数	是否必选	参数类型	描述
version_id	否	String	参数解释： 模型版本ID。 约束限制： 不涉及。 取值范围： 长度为[32,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
name	否	String	参数解释： 模型版本名称，支持模糊查询。 约束限制： 不涉及。 取值范围： 长度为[4,32]的英文字符、数字、中划线(-)和下划线(_)的组合。 默认取值： 不涉及。

请求参数

表 4-269 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

响应参数

状态码：200

表 4-270 响应 Body 参数

参数	参数类型	描述
total	Integer	参数解释： 符合条件的Version总数。 约束限制： 不涉及。 取值范围： [0,10000]。 默认取值： 不涉及。

参数	参数类型	描述
versions	Array of ModelVersionInfo objects	参数解释：列表信息。 约束限制：[0,10000]。

表 4-271 ModelVersionInfo

参数	参数类型	描述
id	String	参数解释：模型版本ID。 约束限制：系统自动生成无法修改，输入不生效。 取值范围：长度为[32,36]的英文字符、数字和中划线(-)的组合。 默认取值：不涉及。
name	String	参数解释：模型版本名称。 约束限制：不涉及。 取值范围：长度为[1,64]的中文、字母、数字、下划线(_)、中划线(-)、半角句号(.)、空格组合。 默认取值：不涉及。
cap_white_list	Array of strings	参数解释：镜像包白名单。 约束限制：[0,100]。
description	String	参数解释：描述信息。 约束限制：不涉及。 取值范围：[0,1024]。 默认取值：不涉及。
config	ModelConfig object	参数解释：模型配置。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
create_time	String	参数解释：创建时间。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
create_user	User object	参数解释：租户信息内容，包含账户名称、账户ID和用户名称、用户ID。 约束限制：不涉及。

表 4-272 ModelConfig

参数	参数类型	描述
llm_model_config	LlmModelConfig object	参数解释: Llm模型配置。 约束限制: 不涉及。

表 4-273 LlmModelConfig

参数	参数类型	描述
base_model_type	String	参数解释: 基模型类型。 约束限制: 不涉及。 取值范围: 请从 列举基模型 接口响应中获取。 默认取值: 不涉及。
model_path	String	参数解释: 模型文件路径。 约束限制: 合法OBS路径。 取值范围: 不涉及。 默认取值: 不涉及。
local_model_path	String	参数解释: 本地模型文件路径。 约束限制: 本地模型文件路径。 取值范围: 不涉及。 默认取值: 不涉及。

表 4-274 User

参数	参数类型	描述
domain_name	String	参数解释: 账户名称。 约束限制: 不涉及。 取值范围: 长度为[0,64]个字符。 默认取值: 不涉及。
domain_id	String	参数解释: 账号ID。 约束限制: 不涉及。 取值范围: 长度为[1,36]的数字、英文字符和中划线(-)的组合。 默认取值: 不涉及。

参数	参数类型	描述
user_name	String	参数解释：用户名称。 约束限制：不涉及。 取值范围：长度为[0,64]个字符。 默认取值：不涉及。
user_id	String	参数解释：用户ID。 约束限制：不涉及。 取值范围：长度为[1,36]的数字、英文字符和中划线(-)的组合。 默认取值：不涉及。

状态码：400

表 4-275 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-276 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404**表 4-277** 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408**表 4-278** 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-279 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

GET https://{endpoint}/v1/workspaces/{workspace_id}/models/{model_id}/versions

响应示例

状态码：200

查询模型版本列表的响应体。

```
{
  "total": 1,
  "versions": [ {
    "id": "0b5633ba2b904511ad514346f4d23d4b",
    "name": "name",
```

```
"cap_white_list" : [ "cap1 cap2 ca* cap*" ],
"description" : "description",
"config" : {
  "llm_model_config" : {
    "base_model_type" : "string",
    "model_path" : "obs://bucketname/objectdir/modelpath/"
  }
},
"create_time" : "2023-05-30T12:24:30.401Z",
"create_user" : {
  "domain_name" : "string",
  "domain_id" : "0b5633ba2b904511ad514346f4d23d4b",
  "user_name" : "string",
  "user_id" : "0b5633ba2b904511ad514346f4d23d4b"
}
}]
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	查询模型版本列表的响应体。
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.7.7 删除未使用的模型定义

功能介绍

清理未使用的模型定义。用户可通过该接口删除模型。此接口为同步接口，无配套使用接口。如果该模型下存在版本正在被使用的情况，则无法删除。需要先删除使用该模型版本的推理服务，才能删除该模型。

URI

DELETE /v1/workspaces/{workspace_id}/models/{model_id}

表 4-280 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： workspaceID。获取方法，请参见 获取workspaceID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
model_id	是	String	参数解释： 推理模型ID。获取方法，请参见 获取模型ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-281 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

响应参数

状态码：200

表 4-282 响应 Body 参数

参数	参数类型	描述
-	AnyType	OK

状态码：400

表 4-283 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-284 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-285 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-286 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-287 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

DELETE https://{endpoint}/v1/workspaces/{workspace_id}/models/{model_id}

响应示例

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	OK
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.8 TMS 管理

4.8.1 查询项目标签

功能介绍

查询项目标签。通过该接口可查询该项目已使用的标签；支持输入：使用预定义标签（为true时可查询预定义的标签）；输出为项目标签。此接口为同步接口，无配套使用接口和特殊场景。

URI

GET /v1/{project_id}/fabric-workspace/tags

表 4-288 路径参数

参数	是否必选	参数类型	描述
project_id	是	String	参数解释： 项目ID。获取方法，请参见 获取项目ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

表 4-289 Query 参数

参数	是否必选	参数类型	描述
use_predefine_tags	是	Boolean	参数解释： 使用预定义标签。 约束限制： 不涉及。 取值范围： true表示使用，false表示不使用。 默认取值： false。
offset	否	Integer	参数解释： 偏移量，表示从此偏移量开始查询。 约束限制： 不涉及。 取值范围： [0,1000000000]。 默认取值： 0。
limit	否	Integer	参数解释： 查询记录数。 约束限制： 不涉及。 取值范围： [1,1000]。 默认取值： 1000。

请求参数

表 4-290 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

响应参数

状态码：200

表 4-291 响应 Header 参数

参数	参数类型	描述
X-request-id	String	参数解释： 请求ID。 约束限制： 不可修改。 取值范围： 不涉及。 默认取值： 不涉及。

表 4-292 响应 Body 参数

参数	参数类型	描述
tags	Array of ResourceTagParam objects	参数解释： 标签列表。 约束限制： [0,100000]。

表 4-293 ResourceTagParam

参数	参数类型	描述
key	String	参数解释： 资源标签的键。 约束限制： 不涉及。 取值范围： 见 查询标签键列表 。 默认取值： 不涉及。

参数	参数类型	描述
values	Array of strings	参数解释：资源标签的值列表。 约束限制：不涉及。

状态码：400

表 4-294 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-295 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-296 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-297 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

```
GET https://{endpoint}/v1/{project_id}/fabric-workspace/tags
```

响应示例

状态码：200

OK

```
{
  "tags" : [ {
    "key" : "tag_key",
    "values" : [ "tag_value" ]
  } ]
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch., cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	OK
400	BadRequest
401	Unauthorized
403	Forbidden
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.8.2 查询资源标签

功能介绍

查询工作空间的标签。用于查询工作空间的标签，返回对应工作空间的标签。此接口为同步接口，无配套使用接口和特殊场景。

URI

GET /v1/{project_id}/fabric-workspace/{workspace_id}/tags

表 4-298 路径参数

参数	是否必选	参数类型	描述
project_id	是	String	参数解释： 项目ID。获取方法，请参见 获取项目ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
workspace_id	是	String	参数解释： 工作空间ID，获取方法，请参见 获取工作空间ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

表 4-299 Query 参数

参数	是否必选	参数类型	描述
offset	否	Integer	参数解释： 偏移量，表示从此偏移量开始查询。 约束限制： 不涉及。 取值范围： [0,10000000000]。 默认取值： 0。
limit	否	Integer	参数解释： 查询记录数。 约束限制： 不涉及。 取值范围： [1,1000]。 默认取值： 1000。

请求参数

表 4-300 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户Token。用于调用接口的认证。获取方法，请参见认证鉴权。 约束限制：不涉及。 取值范围：长度不超过65534个字符。 默认取值：不涉及。

响应参数

状态码：200

表 4-301 响应 Header 参数

参数	参数类型	描述
X-request-id	String	参数解释：请求ID。 约束限制：不可修改。 取值范围：不涉及。 默认取值：不涉及。

表 4-302 响应 Body 参数

参数	参数类型	描述
tags	Array of ResourceTag objects	参数解释：资源标签列表。 约束限制：[0,100000]。
sys_tags	Array of SystemTag objects	参数解释：系统标签列表。 约束限制：[0,1]。

表 4-303 ResourceTag

参数	参数类型	描述
key	String	参数解释：资源标签的键。 约束限制：不涉及。 取值范围：长度1到128的多语言字母、空格、制表符、数字、下划线(_)、点(.)、冒号(:)、斜杠(/)、等号(=)、加号(+)、减号(-)、@符号(@)组合，见查询标签键列表。 默认取值：不涉及。
value	String	参数解释：资源标签的值。 约束限制：不涉及。 取值范围：长度0到255的多语言字母、空格、制表符、数字、下划线(_)、点(.)、冒号(:)、斜杠(/)、等号(=)、加号(+)、减号(-)、@符号(@)组合。 默认取值：不涉及。

表 4-304 SystemTag

参数	参数类型	描述
key	String	参数解释：键。 约束限制：固定值。 取值范围：固定值： _sys_enterprise_project_id。 默认取值：不涉及。
value	String	参数解释：企业项目ID。获取方法，请参见获取企业项目ID。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

状态码：400

表 4-305 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-306 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-307 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-308 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

GET https://{endpoint}/v1/{project_id}/fabric-workspace/{workspace_id}/tags

响应示例

状态码：200

OK

```
{
  "tags" : [ {
    "key" : "key1",
    "value" : "value1"
  } ],
  "sys_tags" : [ {
    "key" : "_sys_enterprise_project_id",
    "value" : "01049549-82cd-4b2b-9733-ddb94350c125"
  } ]
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	OK
400	BadRequest
401	Unauthorized
403	Forbidden
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.8.3 查询资源实例列表

功能介绍

通过标签查询工作空间。输入`without_any_tag=true`查询无标签资源，忽略`tags`字段；`without_any_tag=false`或未提供时，按`tags`、`matches`等条件过滤。`tags`最多含20个键，每键最多20个值，键不可重复，值可为空但结构不可缺失。`matches`支持特定字段的精确或模糊搜索。返回符合过滤条件的工作空间信息。

URI

POST /v1/{project_id}/fabric-workspace/resource-instances/filter

表 4-309 路径参数

参数	是否必选	参数类型	描述
project_id	是	String	参数解释： 项目ID。获取方法，请参见 获取项目ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

表 4-310 Query 参数

参数	是否必选	参数类型	描述
offset	否	Integer	参数解释： 偏移量，表示从此偏移量开始查询。 约束限制： 不涉及。 取值范围： [0,1000000000]。 默认取值： 0。
limit	否	Integer	参数解释： 指定每一页返回的最大条目数。 约束限制： 不涉及。 取值范围： [1,100]。 默认取值： 10。

请求参数

表 4-311 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

表 4-312 请求 Body 参数

参数	是否必选	参数类型	描述
without_any_tag	否	Boolean	参数解释： 不包含任意一个标签，该字段为true时查询所有不带标签的资源，此时忽略“tags”、“tags_any”、“not_tags”、“not_tags_any”字段。 约束限制： 不涉及。 取值范围： true, false。 默认取值： 不涉及。

参数	是否必选	参数类型	描述
tags	否	Array of ResourceTagParam objects	参数解释： 包含标签，最多包含50个key，每个key下面的value最多10个，每个key对应的value可以为空数组但结构体不能缺失。 约束限制： [1,20]。
matches	否	Array of TagMatch objects	参数解释： 搜索字段，key为要匹配的字段，如resource_name等。value为匹配的值。 约束限制： [1,10]。
sys_tags	否	Array of SystemTagParam objects	参数解释： 系统标签列表，目前只包含一个tag结构体。key下面只包含一个value。 约束限制： 1。

表 4-313 ResourceTagParam

参数	是否必选	参数类型	描述
key	是	String	参数解释： 资源标签的键。 约束限制： 不涉及。 取值范围： 见 查询标签键列表 。 默认取值： 不涉及。
values	否	Array of strings	参数解释： 资源标签的值列表。 约束限制： 不涉及。

表 4-314 TagMatch

参数	是否必选	参数类型	描述
key	是	String	参数解释： 键。 约束限制： 不涉及。 取值范围： resource_name。 默认取值： resource_name。
value	否	String	参数解释： 资源标签的值。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

表 4-315 SystemTagParam

参数	是否必选	参数类型	描述
key	是	String	参数解释：键。 约束限制：固定值。 取值范围：固定值： _sys_enterprise_project_id。 默认取值：不涉及。
values	是	Array of strings	参数解释：系统标签的值列表。 约束限制：固定值。

响应参数

状态码：200

表 4-316 响应 Header 参数

参数	参数类型	描述
X-request-id	String	参数解释：请求ID。 约束限制：不可修改。 取值范围：不涉及。 默认取值：不涉及。

表 4-317 响应 Body 参数

参数	参数类型	描述
resources	Array of TagFabricWorkspace objects	参数解释：资源列表。 约束限制：[0,1000]。
total_count	Integer	参数解释：资源总数。 约束限制：不涉及。 取值范围：[0,1000]。 默认取值：不涉及。

表 4-318 TagFabricWorkspace

参数	参数类型	描述
resource_id	String	参数解释： 工作空间的ID。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
resource_detail	Object	参数解释： 资源详情。资源对象，用于扩展。默认为空。 约束限制： 不涉及。
resource_name	String	参数解释： 工作空间名称。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
tags	Array of ResourceTag objects	参数解释： 标签列表，没有标签默认为空数组。 约束限制： [0,100000]。
sys_tags	Array of SystemTag objects	参数解释： 系统标签。 约束限制： 不涉及。

表 4-319 ResourceTag

参数	参数类型	描述
key	String	参数解释： 资源标签的键。 约束限制： 不涉及。 取值范围： 长度1到128的多语言字母、空格、制表符、数字、下划线(_)、点(.)、冒号(:)、斜杠(/)、等号(=)、加号(+)、减号(-)、@符号(@)组合，见 查询标签键列表 。 默认取值： 不涉及。
value	String	参数解释： 资源标签的值。 约束限制： 不涉及。 取值范围： 长度0到255的多语言字母、空格、制表符、数字、下划线(_)、点(.)、冒号(:)、斜杠(/)、等号(=)、加号(+)、减号(-)、@符号(@)组合。 默认取值： 不涉及。

表 4-320 SystemTag

参数	参数类型	描述
key	String	参数解释：键。 约束限制：固定值。 取值范围：固定值： _sys_enterprise_project_id。 默认取值：不涉及。
value	String	参数解释：企业项目ID。获取方法，请参见 获取企业项目ID 。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

状态码：400

表 4-321 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-322 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-323 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-324 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

请求过滤tag_key的标签的资源，具体请求实例如下。

```
POST https://{endpoint}/v1/{project_id}/fabric-workspace/resource-instances/filter
{
  "without_any_tag" : false,
  "tags" : [ {
    "key" : "tag_key",
    "values" : [ "tag_value" ]
  } ],
  "matches" : [ {
    "key" : "resource_name",
    "value" : "tag_value"
  } ],
  "sys_tags" : ""
}
```

响应示例

状态码：200

OK

```
{
  "resources" : [ {
    "resource_id" : "01049549-82cd-4b2b-9733-ddb94350c125",
    "resource_detail" : { },
    "resource_name" : "resource_name",
    "tags" : [ {
      "key" : "key1",
      "value" : "value1"
    } ],
    "sys_tags" : [ {
      "key" : "_sys_enterprise_project_id",
      "value" : "01049549-82cd-4b2b-9733-ddb94350c125"
    } ]
  } ]
}
```

```
    } ],  
    "total_count" : 1  
  }  
}
```

状态码：400

BadRequest

```
{  
  "error_code" : "common.01000001",  
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,  
cause: TypeMismatchException"  
}
```

状态码：401

Unauthorized

```
{  
  "error_code" : "APIG.1002",  
  "error_msg" : "Incorrect token or token resolution failed"  
}
```

状态码：403

Forbidden

```
{  
  "error" : {  
    "code" : "403",  
    "message" : "X-Auth-Token is invalid in the request",  
    "title" : "Forbidden"  
  },  
  "error_code" : 403,  
  "error_msg" : "X-Auth-Token is invalid in the request",  
  "title" : "Forbidden"  
}
```

状态码：408

Request Time-out

```
{  
  "error_code" : "common.00000408",  
  "error_msg" : "timeout exception occurred"  
}
```

状态码：500

InternalServerError

```
{  
  "error_code" : "common.00000500",  
  "error_msg" : "internal error"  
}
```

状态码

状态码	描述
200	OK
400	BadRequest
401	Unauthorized

状态码	描述
403	Forbidden
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.8.4 查询资源实例数量

功能介绍

通过标签查询工作空间数量：*without_any_tag=true*时查询无标签资源，忽略*tags*字段；*false*或未提供时，按*tags*、*matches*等条件过滤。*tags*最多含20个键，每键最多20个值，键不可重复，值可为空但结构不可缺失。*matches*支持特定字段的精确或模糊搜索。返回符合过滤条件的工作空间数量。

URI

POST /v1/{project_id}/fabric-workspace/resource-instances/count

表 4-325 路径参数

参数	是否必选	参数类型	描述
project_id	是	String	参数解释： 项目ID。获取方法，请参见 获取项目ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-326 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户Token。用于调用接口的认证。获取方法，请参见认证鉴权。 约束限制：不涉及。 取值范围：长度不超过65534个字符。 默认取值：不涉及。

表 4-327 请求 Body 参数

参数	是否必选	参数类型	描述
without_any_tag	否	Boolean	参数解释：不包含任意一个标签，该字段为true时查询所有不带标签的资源，此时忽略“tags”、“tags_any”、“not_tags”、“not_tags_any”字段。 约束限制：不涉及。 取值范围：true, false。 默认取值：不涉及。
tags	否	Array of ResourceTagParam objects	参数解释：包含标签，最多包含50个key，每个key下面的value最多10个，每个key对应的value可以为空数组但结构体不能缺失。 约束限制：[1,20]。
matches	否	Array of TagMatch objects	参数解释：搜索字段，key为要匹配的字段，如resource_name等。value为匹配的值。 约束限制：[1,10]。
sys_tags	否	Array of SystemTagParam objects	参数解释：系统标签列表，目前只包含一个tag结构体。key下面只包含一个value。 约束限制：1。

表 4-328 ResourceTagParam

参数	是否必选	参数类型	描述
key	是	String	参数解释：资源标签的键。 约束限制：不涉及。 取值范围：见 查询标签键列表 。 默认取值：不涉及。
values	否	Array of strings	参数解释：资源标签的值列表。 约束限制：不涉及。

表 4-329 TagMatch

参数	是否必选	参数类型	描述
key	是	String	参数解释：键。 约束限制：不涉及。 取值范围：resource_name。 默认取值：resource_name。
value	否	String	参数解释：资源标签的值。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

表 4-330 SystemTagParam

参数	是否必选	参数类型	描述
key	是	String	参数解释：键。 约束限制：固定值。 取值范围：固定值： _sys_enterprise_project_id。 默认取值：不涉及。
values	是	Array of strings	参数解释：系统标签的值列表。 约束限制：固定值。

响应参数

状态码：200

表 4-331 响应 Header 参数

参数	参数类型	描述
X-request-id	String	参数解释： 请求ID，定位辅助信息。 约束限制： 不可修改。 取值范围： 不涉及。 默认取值： 不涉及。

表 4-332 响应 Body 参数

参数	参数类型	描述
total_count	Integer	实例总数。

状态码：400

表 4-333 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释： 错误码。 约束限制： 不涉及。 取值范围： 长度为[8,36]个字符。 默认取值： 不涉及。
error_msg	String	参数解释： 错误描述。 约束限制： 不涉及。 取值范围： 长度为[2,512]个字符。 默认取值： 不涉及。
solution_msg	String	参数解释： 解决方案描述。 约束限制： 不涉及。 取值范围： 长度为[0,4096]个字符。 默认取值： 不涉及。

状态码：401

表 4-334 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-335 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-336 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-337 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

请求统计tag_key的标签的资源，具体请求实例如下。

```
POST https://{endpoint}/v1/{project_id}/fabric-workspace/resource-instances/count
{
```

```
"without_any_tag" : false,
"tags" : [ {
  "key" : "tag_key",
  "values" : [ "tag_value" ]
} ],
"matches" : [ {
  "key" : "resource_name",
  "value" : "tag_value"
} ],
"sys_tags" : ""
}
```

响应示例

状态码：200

OK

```
{
  "total_count" : 1
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code": "common.00000408",
  "error_msg": "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code": "common.00000500",
  "error_msg": "internal error"
}
```

状态码

状态码	描述
200	OK
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.8.5 批量添加资源标签

功能介绍

工作空间批量添加标签。输入：tags标签列表，租户权限时该字段必选。返回值状态码204正常，其他异常。此接口为同步接口，无配套使用接口。创建时，不允许设置重复key数据，如果数据库已存在该key，就覆盖value的值。

URI

POST /v1/{project_id}/fabric-workspace/{workspace_id}/tags/create

表 4-338 路径参数

参数	是否必选	参数类型	描述
project_id	是	String	参数解释： 项目ID。获取方法，请参见 获取项目ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
workspace_id	是	String	参数解释： 工作空间ID。获取方法，请参见 获取工作空间ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-339 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

表 4-340 请求 Body 参数

参数	是否必选	参数类型	描述
tags	否	Array of ResourceTag objects	参数解释： 标签列表。 约束限制： [1,20]。
sys_tags	否	Array of SystemTag objects	参数解释： 系统标签列表。 约束限制： 1。

表 4-341 ResourceTag

参数	是否必选	参数类型	描述
key	是	String	参数解释：资源标签的键。 约束限制：不涉及。 取值范围：长度1到128的多语言字母、空格、制表符、数字、下划线(_)、点(.)、冒号(:)、斜杠(/)、等号(=)、加号(+)、减号(-)、@符号(@)组合，见查询标签键列表。 默认取值：不涉及。
value	否	String	参数解释：资源标签的值。 约束限制：不涉及。 取值范围：长度0到255的多语言字母、空格、制表符、数字、下划线(_)、点(.)、冒号(:)、斜杠(/)、等号(=)、加号(+)、减号(-)、@符号(@)组合。 默认取值：不涉及。

表 4-342 SystemTag

参数	是否必选	参数类型	描述
key	是	String	参数解释：键。 约束限制：固定值。 取值范围：固定值： _sys_enterprise_project_id。 默认取值：不涉及。
value	是	String	参数解释：企业项目ID。获取方法，请参见获取企业项目ID。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

响应参数

状态码：204

表 4-343 响应 Body 参数

参数	参数类型	描述
-	AnyType	OK

状态码：400

表 4-344 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-345 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-346 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-347 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-348 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

请求批量添加资源的标签，具体请求实例如下。

```
POST https://{endpoint}/v1/{project_id}/fabric-workspace/{workspace_id}/tags/create
{
  "tags": [ {
    "key": "key1",
    "value": "value1"
  } ],
  "sys_tags": [ {
    "key": "_sys_enterprise_project_id",
    "value": "01049549-82cd-4b2b-9733-ddb94350c125"
  } ]
}
```

响应示例

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error": {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
204	OK
400	BadRequest
401	Unauthorized

状态码	描述
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.8.6 批量删除资源标签

功能介绍

工作空间批量删除标签。输入：tags标签列表，租户权限时该字段必选。返回值状态码204正常，其他异常。此接口为同步接口，无配套使用接口。删除时，如果删除的标签不存在，默认处理成功，删除时不对标签字符集范围做校验。删除时tags结构体不能缺失，key不能为空，或者空字符串。

URI

POST /v1/{project_id}/fabric-workspace/{workspace_id}/tags/delete

表 4-349 路径参数

参数	是否必选	参数类型	描述
project_id	是	String	参数解释： 项目ID。获取方法，请参见 获取项目ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
workspace_id	是	String	参数解释： 工作空间ID。获取方法，请参见 获取工作空间ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-350 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

表 4-351 请求 Body 参数

参数	是否必选	参数类型	描述
tags	否	Array of DeleteResourceTag objects	参数解释： 批量删除请求体标签列表。 约束限制： [1,20]。
sys_tags	否	Array of DeleteResourceTag objects	参数解释： 批量删除请求体系统标签列表。 约束限制： 1。

表 4-352 DeleteResourceTag

参数	是否必选	参数类型	描述
key	是	String	参数解释： 键。获取方式：标签管理服务TMS/API参考/API说明/资源标签/查询标签键列表。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
value	否	String	参数解释： 值。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

响应参数

状态码：200

表 4-353 响应 Body 参数

参数	参数类型	描述
-	AnyType	OK

状态码：400

表 4-354 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-355 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-356 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-357 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释： 解决方案描述。 约束限制： 不涉及。 取值范围： 长度为[0,4096]个字符。 默认取值： 不涉及。

状态码：500

表 4-358 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释： 错误码。 约束限制： 不涉及。 取值范围： 长度为[8,36]个字符。 默认取值： 不涉及。
error_msg	String	参数解释： 错误描述。 约束限制： 不涉及。 取值范围： 长度为[2,512]个字符。 默认取值： 不涉及。
solution_msg	String	参数解释： 解决方案描述。 约束限制： 不涉及。 取值范围： 长度为[0,4096]个字符。 默认取值： 不涉及。

请求示例

请求批量删除资源的标签，具体请求实例如下。

```
POST https://{endpoint}/v1/{project_id}/fabric-workspace/{workspace_id}/tags/delete
{
  "tags": [ {
    "key": "key1",
    "value": "value1"
  } ],
  "sys_tags": [ {
    "key": "key1",
    "value": "value1"
  } ]
}
```

响应示例

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error": {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	OK
400	BadRequest
401	Unauthorized

状态码	描述
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.9 消息管理

4.9.1 创建消息通知策略

功能介绍

创建消息通知策略。用户通过调用该接口创建一个消息通知策略。任务执行完成时会
将结果根据匹配到的消息通知策略发送消息。支持传入SMN主题以及任务名称匹配正
则表达式。此接口为同步接口，无配套使用接口和特殊场景。

URI

POST /v1/workspaces/{workspace_id}/messages

表 4-359 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间ID，获取方法，请参见 获取工作空间ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-360 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户Token。用于调用接口的认证。获取方法，请参见认证鉴权。 约束限制：不涉及。 取值范围：长度不超过65534个字符。 默认取值：不涉及。

表 4-361 请求 Body 参数

参数	是否必选	参数类型	描述
message_notification_policy_list	是	Array of CreateMessageNotificationPolicy objects	参数解释：消息通知策略。 约束限制：不涉及。 取值范围：[0,1000]。 默认取值：不涉及。

表 4-362 CreateMessageNotificationPolicy

参数	是否必选	参数类型	描述
message_type	是	String	参数解释：消息类型。 约束限制：不涉及。 取值范围： job：任务执行结果。 默认取值：job。
name_pattern	是	String	参数解释：名称样式，用来匹配消息类型中所有符合该样式的信息。例如：message_type设置为job，name_pattern设置为ray_job*，表示匹配到所有以“ray_job”开头的job发出的消息。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

参数	是否必选	参数类型	描述
notification_types	是	Array of strings	参数解释： 通知类型。 约束限制： SUCCESS:成功通知；FAILED：失败通知。
topic_urn	是	String	参数解释： 消息通知主题。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

响应参数

状态码：200

表 4-363 响应 Body 参数

参数	参数类型	描述
id	String	参数解释： 消息通知策略ID。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
message_type	String	参数解释： 消息类型。 约束限制： 不涉及。 取值范围： • job：任务执行结果。 默认取值： job。
name_pattern	String	参数解释： 名称样式，用来匹配消息类型中所有符合该样式的消息。例如：message_type设置为job，name_pattern设置为ray_job*，表示匹配到所有以"ray_job"开头的job发出的消息 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
notification_types	Array of strings	参数解释： 通知类型。SUCCESS：成功通知；FAILED：失败通知。 约束限制： [1,2]。

参数	参数类型	描述
topic_urn	String	参数解释：消息通知主题。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

状态码：400

表 4-364 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-365 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-366 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-367 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-368 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

创建消息通知策略。具体请求样例如下。

```
POST https://{endpoint}/v1/workspaces/{workspace_id}/messages
{
  "message_notification_policy_list" : [ {
    "message_type" : "job",
    "name_pattern" : "ray_job*",
    "notification_types" : [ "SUCCESS", "FAILED" ],
    "topic_urn" : "smn:topic:xxxxxxx"
  } ]
}
```

响应示例

状态码：200

消息通知策略响应体

```
{
  "id" : "ray_job_123456789",
}
```

```
{
  "message_type" : "job",
  "name_pattern" : "ray_job*",
  "notification_types" : [ "SUCCESS", "FAILED" ],
  "topic_urn" : "smn:topic:xxxxxxx"
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	消息通知策略响应体
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.9.2 查询消息通知策略列表

功能介绍

查询消息通知策略。用户通过调用该接口查询已创建的消息通知策略。支持分页查询，根据传入的工作空间ID，消息匹配样式来查询。此接口为同步接口，无配套使用接口和特殊场景。

URI

GET /v1/workspaces/{workspace_id}/messages

表 4-369 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间ID，获取方法，请参见 获取工作空间ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

表 4-370 Query 参数

参数	是否必选	参数类型	描述
offset	否	Integer	参数解释： 偏移量，表示从此偏移量开始查询。 约束限制： 不涉及。 取值范围： [0,1000000000]。 默认取值： 0。
limit	否	Integer	参数解释： 指定每一页返回的最大条目数。 约束限制： 不涉及。 取值范围： [1,100]。 默认取值： 10。
message_type	否	String	参数解释： 消息类型。 约束限制： 不可修改。 取值范围： job：任务执行结果消息。 默认取值： 不涉及。
name_pattern	否	String	参数解释： 名称样式。 约束限制： 支持模糊匹配，区分大小写。 取值范围： 不涉及。 默认取值： 不涉及。
notify_type	否	String	参数解释： 消息类型。 约束限制： 不涉及。 取值范围： ● SUCCESS：成功通知； ● FAILED：失败通知。 默认取值： 不涉及。

请求参数

表 4-371 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

响应参数

状态码：200

表 4-372 响应 Body 参数

参数	参数类型	描述
total	Integer	参数解释： 符合条件的消息通知策略总数。 约束限制： 不涉及。 取值范围： [0,10000]。 默认取值： 不涉及。
policies	Array of MessageNotificationPolicy objects	参数解释： 消息通知策略列表信息。 约束限制： [0,10000]。

表 4-373 MessageNotificationPolicy

参数	参数类型	描述
id	String	参数解释： 消息通知策略ID。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
message_type	String	参数解释： 消息类型。 约束限制： 不涉及。 取值范围： job: 任务执行结果。 默认取值： job。

参数	参数类型	描述
name_pattern	String	参数解释： 名称样式，用来匹配消息类型中所有符合该样式的消息。例如：message_type设置为job，name_pattern设置为ray_job*，表示匹配到所有以"ray_job"开头的job发出的消息 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
notification_types	Array of strings	参数解释： 通知类型。SUCCESS：成功通知；FAILED：失败通知。 约束限制： [1,2]。
topic_urn	String	参数解释： 消息通知主题。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

状态码：400

表 4-374 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释： 错误码。 约束限制： 不涉及。 取值范围： 长度为[8,36]个字符。 默认取值： 不涉及。
error_msg	String	参数解释： 错误描述。 约束限制： 不涉及。 取值范围： 长度为[2,512]个字符。 默认取值： 不涉及。
solution_msg	String	参数解释： 解决方案描述。 约束限制： 不涉及。 取值范围： 长度为[0,4096]个字符。 默认取值： 不涉及。

状态码：401

表 4-375 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-376 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-377 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-378 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

GET https://{endpoint}/v1/workspaces/{workspace_id}/messages

响应示例

状态码：200

查询消息通知策略列表响应体。

```
{
  "total" : 1,
  "policies" : [ {
    "id" : "ray_job_123456789",
    "message_type" : "job",
    "name_pattern" : "ray_job*",
    "notification_types" : [ "SUCCESS", "FAILED" ],
    "topic_urn" : "smn:topic:xxxxxxx"
  } ]
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	查询消息通知策略列表响应体。
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.9.3 删除消息通知策略

功能介绍

删除消息通知策略。用户通过调用该接口删除已创建的消息通知策略。此接口为同步接口，无配套使用接口和特殊场景。

URI

DELETE /v1/workspaces/{workspace_id}/messages/{message_policy_id}

表 4-379 路径参数

参数	是否必选	参数类型	描述
message_policy_id	是	String	参数解释： 消息通知策略ID。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间ID，获取方法，请参见 获取工作空间ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-380 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制： 不涉及。 取值范围： 长度不超过65534个字符。 默认取值： 不涉及。

响应参数

状态码：200

OK

状态码：400

表 4-381 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释： 错误码。 约束限制： 不涉及。 取值范围： 长度为[8,36]个字符。 默认取值： 不涉及。
error_msg	String	参数解释： 错误描述。 约束限制： 不涉及。 取值范围： 长度为[2,512]个字符。 默认取值： 不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-382 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-383 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-384 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-385 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

```
GET https://{endpoint}/v1/workspaces/{workspace_id}/messages/{message_policy_id}
```

响应示例

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
}
```

```
{  
  "error_msg": "timeout exception occurred"  
}
```

状态码：500

InternalServerError

```
{  
  "error_code": "common.00000500",  
  "error_msg": "internal error"  
}
```

状态码

状态码	描述
200	OK
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.10 监控管理

4.10.1 更新 AOM 监控采集配置

功能介绍

更新工作空间下的AOM监控采集配置。此接口为同步接口，无配套使用接口和特殊场景。

URI

PUT /v1/workspaces/{workspace_id}/metrics

表 4-386 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释：工作空间ID，获取方法，请参见获取工作空间ID。 约束限制：不涉及。 取值范围：长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值：不涉及。

请求参数

表 4-387 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户Token。用于调用接口的认证。获取方法，请参见认证鉴权。 约束限制：不涉及。 取值范围：长度不超过65534个字符。 默认取值：不涉及。

表 4-388 请求 Body 参数

参数	是否必选	参数类型	描述
enable	是	Boolean	参数解释：是否开启AOM监控采集。 约束限制：不涉及。 取值范围：false为不开启监控采集，true开启监控采集。 默认取值：false。

响应参数

状态码：200

表 4-389 响应 Body 参数

参数	参数类型	描述
-	AnyType	OK

状态码：400

表 4-390 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-391 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-392 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-393 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释： 解决方案描述。 约束限制： 不涉及。 取值范围： 长度为[0,4096]个字符。 默认取值： 不涉及。

请求示例

更新工作空间下的AOM监控采集配置为false。

```
PUT https://{endpoint}/v1/workspaces/{workspace_id}/metrics
{
  "enable" : false
}
```

响应示例

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch., cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	OK
400	BadRequest
401	Unauthorized
403	Forbidden
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.11 接入客户端

4.11.1 创建 vpcep 客户端

功能介绍

创建vpcep客户端。VPC打通APIG，使用户可以通过内网域名调用到DataArtsFabric接口。此接口为同步接口，无配套使用接口和特殊场景。

URI

POST /v1/access-clients

请求参数

表 4-394 请求 Body 参数

参数	是否必选	参数类型	描述
name	否	String	参数解释： 客户端名称。 约束限制： 不涉及。 取值范围： 长度为[4,32]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
vpc_id	否	String	参数解释： 虚拟私有云id。 约束限制： 不涉及。 取值范围： 长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
subnet_id	否	String	参数解释： 子网ID。创建子网时由系统生成。 约束限制： 不涉及。 取值范围： 长度为[36]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
subnet_ids	否	Array of strings	参数解释： 所属子网组。 约束限制： 不涉及。 取值范围： [0,100]。 默认取值： 不涉及。
access_connections	否	Array of AccessConnectionRequestBody objects	接入连接列表，当接入模式为CUSTOM模式时生效。

表 4-395 AccessConnectionRequestBody

参数	是否必选	参数类型	描述
vpcep_id	否	String	参数解释： 虚拟私有云终端节点ID。在“接入管理-创建客户端-前往VPC创建-VPC终端节点”创建和查看。 约束限制： 不涉及。 取值范围： [0,64]。 默认取值： 不涉及。
vpcep_service_name	否	String	参数解释： 终端节点服务名称。 约束限制： 不涉及。 取值范围： [0,64]。 默认取值： 不涉及。
domain	否	String	参数解释： 接入域名，通过IP接入访问Lakeformation API时，需在请求头中添加HOST参数并传入该域名。 约束限制： 不涉及。 取值范围： [0,100]。 默认取值： 不涉及。

响应参数

状态码：200

表 4-396 响应 Body 参数

参数	参数类型	描述
id	String	参数解释： id。 约束限制： 不涉及。 取值范围： 长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
name	String	参数解释： 名称。 约束限制： 不涉及。 取值范围： 长度为[1,64]的中文、字母、数字、下划线(_)、中划线(-)、点(.)、空格的组合。 默认取值： 不涉及。

参数	参数类型	描述
status	String	参数解释： 接入状态。 约束限制： 不涉及。 取值范围： • CREATING：创建中； • RUNNING：运行中； • DELETING：删除中； • DELETED：已删除； 默认取值： 不涉及。
vpc_id	String	参数解释： 服务接入客户端的vpcID。 约束限制： 不涉及。 取值范围： 长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
subnet_ids	Array of strings	参数解释： 服务接入客户端的子网ID，多个子网用“,”连接。 约束限制： 不涉及。 取值范围： [0,100]。 默认取值： 不涉及。
access_connections	Array of AccessConnectionInfo objects	参数解释： 接入连接列表。 约束限制： 不涉及。 取值范围： [0,100]。 默认取值： 不涉及
create_time	String	参数解释： 创建时间。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

表 4-397 AccessConnectionInfo

参数	参数类型	描述
vpcep_id	String	参数解释： vpcep id。 约束限制： 不涉及。 取值范围： 长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及

参数	参数类型	描述
subnet_id	String	参数解释: subnet_id id。 约束限制: 不涉及。 取值范围: 长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值: 不涉及
domain	String	参数解释: 接入域名, 通过IP接入访问 DataArtsFabric API时, 需在请求头中添加HOST参数并传入该域名。 约束限制: 不涉及。 取值范围: 长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值: 不涉及
ip	String	参数解释: ipv4地址。 约束限制: ipv4地址。 取值范围: 不涉及。 默认取值: 不涉及
owner	String	参数解释: 拥有者。 约束限制: 不涉及。 取值范围: 长度[1,64]。 默认取值: 不涉及。

状态码: 400

表 4-398 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释: 错误码。 约束限制: 不涉及。 取值范围: 长度为[8,36]个字符。 默认取值: 不涉及。
error_msg	String	参数解释: 错误描述。 约束限制: 不涉及。 取值范围: 长度为[2,512]个字符。 默认取值: 不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-399 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-400 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-401 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-402 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释： 解决方案描述。 约束限制： 不涉及。 取值范围： 长度为[0,4096]个字符。 默认取值： 不涉及。

请求示例

创建vpcep客户端。具体请求配置如样例所示。

```
POST https://{endpoint}/v1/access-clients

{
  "name": "zzh_test001",
  "vpc_id": "050e154d-9954-4929-b6d3-12af7c9213ac",
  "subnet_id": "string",
  "subnet_ids": [ "0d86d58d-d864-47ce-8dbd-5487ff1d59ba" ],
  "access_connections": [ {
    "vpcep_id": "string",
    "vpcep_service_name": "string",
    "domain": "string"
  } ]
}
```

响应示例

状态码：200

创建接入客户端的响应体。

```
{
  "id": "f7670aea-36a1-4bdf-ae99-963c33324370",
  "name": "name",
  "status": "RUNNING",
  "vpc_id": "f18a378a-f26b-461d-985f-3f904e4cc544",
  "subnet_ids": [ "1e9a3f25-0f1f-40c4-8ef9-e4a8dd4d1e53" ],
  "access_connections": [ {
    "vpcep_id": "1851896e-3731-4072-8b4a-33e38ed2207b",
    "domain": "1e9a3f25-0f1f-40c4-8ef9-e4a8dd4d1e53",
    "ip": "192.168.0.57",
    "owner": "0676610f3a0a4c2c80c50bea7ddf18c1"
  } ],
  "create_time": "2025-01-08T01:53:30.000+00:00"
}
```

状态码：400

BadRequest

```
{
  "error_code": "common.01000001",
  "error_msg": "failed to read http request, please check your input, code: 400, reason: Type mismatch., cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code": "APIG.1002",
}
```

```
{  "error_msg": "Incorrect token or token resolution failed"}
```

状态码：403

Forbidden

```
{  "error": {    "code": "403",    "message": "X-Auth-Token is invalid in the request",    "title": "Forbidden"  },  "error_code": 403,  "error_msg": "X-Auth-Token is invalid in the request",  "title": "Forbidden"}
```

状态码：404

NotFound

```
{  "error_code": "common.01000001",  "error_msg": "response status exception, code: 404"}
```

状态码：408

Request Time-out

```
{  "error_code": "common.00000408",  "error_msg": "timeout exception occurred"}
```

状态码：500

InternalServerError

```
{  "error_code": "common.00000500",  "error_msg": "internal error"}
```

状态码

状态码	描述
200	创建接入客户端的响应体。
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.11.2 查询 vpcpep 客户端列表

功能介绍

查询接入端列表。用户可以查询到已创建的接入端列表，获取VPCEP内网域名，进行使用。此接口为同步接口，无配套使用接口和特殊场景。

URI

GET /v1/access-clients

表 4-403 Query 参数

参数	是否必选	参数类型	描述
offset	否	Integer	参数解释： 偏移量，表示从此偏移量开始查询。 约束限制： 不涉及。 取值范围： [0,10000000000]。 默认取值： 0。
limit	否	Integer	参数解释： 指定每一页返回的最大条目数。 约束限制： 不涉及。 取值范围： [1,100]。 默认取值： 10。
name	否	String	参数解释： 通过vpcep client名称搜索vpcep客户端，支持模糊搜索。 约束限制： 不涉及。 取值范围： 长度为[1,64]的英文字符、中文字符、数字、中划线(-)和下划线(_)的组合。 默认取值： 不涉及。
id	否	String	参数解释： 通过vpcep client ID搜索客户端。 约束限制： 不涉及。 取值范围： 长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-404 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户Token。用于调用接口的认证。获取方法，请参见认证鉴权。 约束限制：不涉及。 取值范围：长度不超过65534个字符。 默认取值：不涉及。

响应参数

状态码：200

表 4-405 响应 Body 参数

参数	参数类型	描述
access_client_infos	Array of AccessClientInfo objects	参数解释：vpcepClient信息。 约束限制：不涉及。 取值范围：[0,100]。 默认取值：不涉及。
total	Integer	参数解释：记录数。 约束限制：不涉及。 取值范围：[0,100]。 默认取值：不涉及。

表 4-406 AccessClientInfo

参数	参数类型	描述
id	String	参数解释：id。 约束限制：不涉及。 取值范围：长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值：不涉及。

参数	参数类型	描述
name	String	参数解释：名称。 约束限制：不涉及。 取值范围：长度为[1,64]的中文、字母、数字、下划线(_)、中划线(-)、点(.)、空格的组合。 默认取值：不涉及。
status	String	参数解释：接入状态。 约束限制：不涉及。 取值范围： • CREATING：创建中； • RUNNING：运行中； • DELETING：删除中； • DELETED：已删除； 默认取值：不涉及。
vpc_id	String	参数解释：服务接入客户端的vpcID。 约束限制：不涉及。 取值范围：长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值：不涉及。
subnet_ids	Array of strings	参数解释：服务接入客户端的子网ID，多个子网用“,”连接。 约束限制：不涉及。 取值范围：[0,100]。 默认取值：不涉及。
access_connections	Array of AccessConnections objects	参数解释：接入连接列表。 约束限制：不涉及。 取值范围：[0,100]。 默认取值：不涉及
create_time	String	参数解释：创建时间。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

表 4-407 AccessConnectionInfo

参数	参数类型	描述
vpcep_id	String	参数解释： vpcep id。 约束限制： 不涉及。 取值范围： 长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及
subnet_id	String	参数解释： subnet_id id。 约束限制： 不涉及。 取值范围： 长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及
domain	String	参数解释： 接入域名，通过IP接入访问DataArtsFabric API时，需在请求头中添加HOST参数并传入该域名。 约束限制： 不涉及。 取值范围： 长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及
ip	String	参数解释： ipv4地址。 约束限制： ipv4地址。 取值范围： 不涉及。 默认取值： 不涉及
owner	String	参数解释： 拥有者。 约束限制： 不涉及。 取值范围： 长度[1,64]。 默认取值： 不涉及。

状态码：400

表 4-408 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释： 错误码。 约束限制： 不涉及。 取值范围： 长度为[8,36]个字符。 默认取值： 不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401**表 4-409** 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404**表 4-410** 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-411 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-412 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

无

响应示例

状态码：200

查询vpcep client列表响应体。

```
{
  "total": 1,
  "access_client_infos": [ {
    "id": "f7670aea-36a1-4bdf-ae99-963c33324370",
    "name": "name",
    "status": "RUNNING",
    "vpc_id": "f18a378a-f26b-461d-985f-3f904e4cc544",
    "subnet_ids": [ "1e9a3f25-0f1f-40c4-8ef9-e4a8dd4d1e53" ],
    "access_connections": [ {
      "vpcep_id": "1851896e-3731-4072-8b4a-33e38ed2207b",
      "domain": "1e9a3f25-0f1f-40c4-8ef9-e4a8dd4d1e53",
      "ip": "192.168.0.57",
      "owner": "0676610f3a0a4c2c80c50bea7ddf18c1"
    } ],
    "create_time": "2025-01-08T01:53:30.000+00:00"
  } ]
}
```

状态码：400

BadRequest

```
{
  "error_code": "common.01000001",
  "error_msg": "failed to read http request, please check your input, code: 400, reason: Type mismatch., cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code": "APIG.1002",
  "error_msg": "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error": {
    "code": "403",
    "message": "X-Auth-Token is invalid in the request",
    "title": "Forbidden"
  },
  "error_code": 403,
  "error_msg": "X-Auth-Token is invalid in the request",
  "title": "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code": "common.01000001",
  "error_msg": "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code": "common.00000408",
  "error_msg": "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code": "common.00000500",
  "error_msg": "internal error"
}
```

状态码

状态码	描述
200	查询vpcep client列表响应体。
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.11.3 查询 vpcep 客户端详情

功能介绍

查询接入端详情。用户可以查看接入端的详情。此接口为同步接口，无配套使用接口和特殊场景。

URI

GET /v1/access-clients/{client_id}

表 4-413 路径参数

参数	是否必选	参数类型	描述
client_id	是	String	参数解释：vpcep client ID。 约束限制：不涉及。 取值范围：长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值：不涉及。

请求参数

表 4-414 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制：不涉及。 取值范围：长度不超过65534个字符。 默认取值：不涉及。

响应参数

状态码：200

表 4-415 响应 Body 参数

参数	参数类型	描述
id	String	参数解释： id。 约束限制： 不涉及。 取值范围： 长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
name	String	参数解释： 名称。 约束限制： 不涉及。 取值范围： 长度为[1,64]的中文、字母、数字、下划线(_)、中划线(-)、点(.)、空格的组合。 默认取值： 不涉及。
status	String	参数解释： 接入状态。 约束限制： 不涉及。 取值范围： • CREATING：创建中； • RUNNING：运行中； • DELETING：删除中； • DELETED：已删除； 默认取值： 不涉及。
vpc_id	String	参数解释： 服务接入客户端的vpcID。 约束限制： 不涉及。 取值范围： 长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及。
subnet_ids	Array of strings	参数解释： 服务接入客户端的子网ID，多个子网用“,”连接。 约束限制： 不涉及。 取值范围： [0,100]。 默认取值： 不涉及。
access_connections	Array of AccessConnectionInfo objects	参数解释： 接入连接列表。 约束限制： 不涉及。 取值范围： [0,100]。 默认取值： 不涉及
create_time	String	参数解释： 创建时间。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

表 4-416 AccessConnectionInfo

参数	参数类型	描述
vpcep_id	String	参数解释： vpcep id。 约束限制： 不涉及。 取值范围： 长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及
subnet_id	String	参数解释： subnet_id id。 约束限制： 不涉及。 取值范围： 长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及
domain	String	参数解释： 接入域名，通过IP接入访问 DataArtsFabric API时，需在请求头中添加HOST参数并传入该域名。 约束限制： 不涉及。 取值范围： 长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值： 不涉及
ip	String	参数解释： ipv4地址。 约束限制： ipv4地址。 取值范围： 不涉及。 默认取值： 不涉及
owner	String	参数解释： 拥有者。 约束限制： 不涉及。 取值范围： 长度[1,64]。 默认取值： 不涉及。

状态码：400

表 4-417 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释： 错误码。 约束限制： 不涉及。 取值范围： 长度为[8,36]个字符。 默认取值： 不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401**表 4-418** 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404**表 4-419** 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-420 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-421 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

无

响应示例

状态码：200

查询vpcep client详情响应体。

```
{
  "id": "f7670aea-36a1-4bdf-ae99-963c33324370",
  "name": "name",
  "status": "RUNNING",
  "vpc_id": "f18a378a-f26b-461d-985f-3f904e4cc544",
  "subnet_ids": [ "1e9a3f25-0f1f-40c4-8ef9-e4a8dd4d1e53" ],
  "access_connections": [ {
    "vpcep_id": "1851896e-3731-4072-8b4a-33e38ed2207b",
    "domain": "1e9a3f25-0f1f-40c4-8ef9-e4a8dd4d1e53",
    "ip": "192.168.0.57",
    "owner": "0676610f3a0a4c2c80c50bea7ddf18c1"
  } ],
  "create_time": "2025-01-08T01:53:30.000+00:00"
}
```

状态码：400

BadRequest

```
{
  "error_code": "common.01000001",
  "error_msg": "failed to read http request, please check your input, code: 400, reason: Type mismatch., cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code": "APIG.1002",
  "error_msg": "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error": {
    "code": "403",
    "message": "X-Auth-Token is invalid in the request",
    "title": "Forbidden"
  },
  "error_code": 403,
  "error_msg": "X-Auth-Token is invalid in the request",
  "title": "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code": "common.01000001",
  "error_msg": "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code": "common.00000408",
  "error_msg": "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code": "common.00000500",
  "error_msg": "internal error"
}
```

状态码

状态码	描述
200	查询vpcep client详情响应体。
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.11.4 删除 vpcep 客户端

功能介绍

删除接入客户端。用户可以删除不再使用的接入端。此接口为同步接口，无配套使用接口和特殊场景。

URI

DELETE /v1/access-clients/{client_id}

表 4-422 路径参数

参数	是否必选	参数类型	描述
client_id	是	String	参数解释：vpcep client ID。 约束限制：不涉及。 取值范围：长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值：不涉及。

请求参数

表 4-423 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制：不涉及。 取值范围：长度不超过65534个字符。 默认取值：不涉及。

响应参数

状态码：200

删除客户端响应

状态码：400

表 4-424 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-425 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-426 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-427 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-428 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

无

响应示例

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
}
```

```
"title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	删除客户端响应
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.11.5 更新 vpcep 客户端

功能介绍

更新接入客户端。用户可以修改接入客户端的名字。此接口为同步接口，无配套使用接口和特殊场景。

URI

PUT /v1/access-clients/{client_id}

表 4-429 路径参数

参数	是否必选	参数类型	描述
client_id	是	String	参数解释：vpcep client ID。 约束限制：不涉及。 取值范围：长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值：不涉及。

请求参数

表 4-430 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户Token。用于调用接口的认证。获取方法，请参见 认证鉴权 。 约束限制：不涉及。 取值范围：长度不超过65534个字符。 默认取值：不涉及。

表 4-431 请求 Body 参数

参数	是否必选	参数类型	描述
name	否	String	参数解释：vpcep客户端名称。 约束限制：不涉及。 取值范围：[1,64]。 默认取值：不涉及。

响应参数

状态码：200

OK

状态码：400

表 4-432 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：401

表 4-433 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：404

表 4-434 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：408

表 4-435 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

状态码：500

表 4-436 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]个字符。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,512]个字符。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度为[0,4096]个字符。 默认取值：不涉及。

请求示例

更新接入客户端。修改接入客户端的名字为zzh_test1122。

```
PUT https://{endpoint}/v1/access-clients/{client_id}
{
  "name" : "zzh_test1122"
}
```

响应示例

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch., cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
  }
}
```

```
{
  "message": "X-Auth-Token is invalid in the request",
  "title": "Forbidden"
},
{
  "error_code": 403,
  "error_msg": "X-Auth-Token is invalid in the request",
  "title": "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code": "common.01000001",
  "error_msg": "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code": "common.00000408",
  "error_msg": "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code": "common.00000500",
  "error_msg": "internal error"
}
```

状态码

状态码	描述
200	OK
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.12 Service 实例管理

4.12.1 列举已部署的服务实例

功能介绍

列举已部署的服务实例列表。用户可通过该接口列举已部署的服务实例列表，接口支持分页查询、支持通过名称、服务实例ID、端点ID、可见性查询。支持排序。返回为符合过滤条件的服务实例列表。此接口为同步接口，无配套使用接口。支持查询公共的服务实例。

URI

GET /v1/workspaces/{workspace_id}/services/instances

表 4-437 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。

表 4-438 Query 参数

参数	是否必选	参数类型	描述
id	否	String	参数解释： 通过服务实例id检索。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。
name	否	String	参数解释： 通过名字搜索服务实例，支持模糊查询。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。
limit	否	Integer	参数解释： 指定每一页返回的最大条目数。 约束限制： 不涉及。 取值范围： [1,100]。 默认取值： 10。

参数	是否必选	参数类型	描述
offset	否	Integer	参数解释： 偏移量，表示从此偏移量开始查询。 约束限制： 不涉及。 取值范围： [0,1000000000]。 默认取值： 0。
endpoint_id	否	String	参数解释： 通过ID检索Endpoint的参数。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。
source_id	否	String	参数解释： Service ID或者Model ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。
version_id	否	String	参数解释： Service version ID或者Model version ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。
type	否	String	参数解释： Service的类型。 约束限制： 不涉及。 取值范围： LLM_MODEL：大语言模型，PGSQL_SERVICE：已废弃。 默认取值： 不涉及。

参数	是否必选	参数类型	描述
visibility	否	String	参数解释： 可见性检索的参数。 约束限制： 不涉及。 取值范围： - PRIVATE: 私有，用户自己创建的； - PUBLIC: 公共，查询所有公共的，包括其他用户创建的。 默认取值： 为空，不填表示不限制，则查出当前用户下的，包括PRIVATE和PUBLIC，不包括其他用户创建的。
sort_by	否	String	参数解释： 根据字段排序。 约束限制： 不涉及。 取值范围： CREATE_TIME: 创建时间。 默认取值： 不涉及。
order_by	否	String	参数解释： 排序方式。 约束限制： 不涉及。 取值范围： ASC: 正序排序；DESC: 倒序排序。 默认取值： DESC。

请求参数

表 4-439 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户token。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

响应参数

状态码：200

表 4-440 响应 Body 参数

参数	参数类型	描述
total	Integer	参数解释： 符合条件的service Instance总数。 约束限制： 不涉及。 取值范围： [0,10000]。 默认取值： 不涉及。
service_instances	Array of ServiceInstanceBriefInfo objects	参数解释： 符合条件的service Instance列表。 约束限制： [0,1000]。

表 4-441 ServiceInstanceBriefInfo

参数	参数类型	描述
visibility	String	参数解释： 可见性。 约束限制： 不涉及。 取值范围： - PRIVATE: 私有 - PUBLIC: 公共 默认取值： PRIVATE。
source	SourceRef object	参数解释： 来源信息，Service信息或者模型信息。 约束限制： 不涉及。
id	String	参数解释： 服务实例的ID。获取方法，请参见 获取推理服务实例ID 。 约束限制： 不涉及。 取值范围： 长度为[32,36]的中文、字母、数字、中划线(-)的组合。 默认取值： 不涉及。
name	String	参数解释： 一个服务实例的名称。 约束限制： 不涉及。 取值范围： 长度为[1,64]的中文、字母、数字、下划线(_)、中划线(-)、点、空格 的组合。 默认取值： 不涉及。

参数	参数类型	描述
description	String	参数解释：描述信息。 约束限制：不涉及。 取值范围：长度为[0,1024]的除了“!”、“<”、“>”、“=”、“&”、“”、“_”之外的任意字符的组合。 默认取值：不涉及。
endpoint_id	String	参数解释：端点ID。 约束限制：不涉及。 取值范围：长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值：不涉及。
status	String	参数解释：运行状态。 约束限制：不涉及。 取值范围： • PENDING：待处理 • CREATING：创建中 • RUNNING：运行中 • UPDATING：更新中 • SUCCEEDED：成功 • FAILED：失败 • STOPPING：停止中 • STOPPED：已停止 • DELETING：删除中 • DELETED：已删除 • INACTIVE：待激活 • TRIAL_ENDED：已结束 默认取值：不涉及。
create_time	String	参数解释：创建时间。 约束限制：不涉及。
update_time	String	参数解释：更新时间。 约束限制：不涉及。
duration	Long	参数解释：运行时长。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

参数	参数类型	描述
create_user	User object	参数解释：租户信息，包含账户名称和用户名称。 约束限制：不涉及。
type	ServiceType object	参数解释：Service的类型。 约束限制：不涉及。 取值范围： - PGSQL_SERVICE：已废弃 - LLM_MODEL：大语言模型 默认取值：不涉及。
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度为[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度为[2,4096]。 默认取值：不涉及。
solution	String	参数解释：解决方案。 约束限制：不涉及。 取值范围：长度为[0,4096]。 默认取值：不涉及。
config	AppInstanceConfig object	参数解释：启动App实例时输入的配置。 约束限制：不涉及。
content_moderation_switch	String	参数解释：模型内容审核开关。 约束限制：不涉及。 取值范围： - ON：开启内容审核； - OFF：关闭内容审核。 默认取值：不涉及。
content_output_moderation_switch	String	参数解释：模型内容审核开关。 约束限制：不涉及。 取值范围： - ON：开启内容审核； - OFF：关闭内容审核。 默认取值：不涉及。

表 4-442 SourceRef

参数	参数类型	描述
id	String	参数解释： 来源ID。有不同的类型，对于推理服务实例，则为模型的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字母、数字、中划线(-)的组合。 默认取值： 不涉及。
version_id	String	参数解释： 版本ID。有不同的类型，对于推理服务实例，则为模型版本ID 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字母、数字、中划线(-)的组合。 默认取值： 不涉及。

表 4-443 User

参数	参数类型	描述
domain_name	String	参数解释： 账户名称。 约束限制： 不涉及。 取值范围： [0,64]。 默认取值： 不涉及。
user_name	String	参数解释： 用户名称。 约束限制： 不涉及。 取值范围： [0,64]。 默认取值： 不涉及。

表 4-444 ServiceType

参数	参数类型	描述
-	String	参数解释： Service的类型。 约束限制： 不涉及。 取值范围： - PGSQL_SERVICE：已废弃 - LLM_MODEL：大语言模型 默认取值： 不涉及。

表 4-445 ApplInstanceConfig

参数	参数类型	描述
model_instance_config	ModelServiceInstanceConfig object	参数解释： 模型服务实例的配置，在DeployService中使用。 约束限制： 不涉及。

表 4-446 ModelServiceInstanceConfig

参数	参数类型	描述
resource	ResourceDemand object	参数解释： 资源需求量配置。 约束限制： 不涉及。

表 4-447 ResourceDemand

参数	参数类型	描述
min	Integer	参数解释： 最小数。 约束限制： 不涉及。 取值范围： [1,1000]。 默认取值： 不涉及。
max	Integer	参数解释： 最大数。 约束限制： 不涉及。 取值范围： [1,1000]。 默认取值： 不涉及。
spec_code	String	参数解释： 资源规格，从规格列表查询获取。 约束限制： 不涉及。 取值范围： 长度为[1,64]。 默认取值： 不涉及。

状态码：400

表 4-448 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：401

表 4-449 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：404

表 4-450 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：408

表 4-451 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：500

表 4-452 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

请求示例

无

响应示例

状态码：200

查询service instance列表的响应体。

```
{
  "total": 1,
  "service_instances": [ {
    "visibility": "PRIVATE",
    "source": {
      "id": "0b5633ba2b904511ad514346f4d23d4b",
      "version_id": "0b5633ba2b904511ad514346f4d23d4b"
    },
    "id": "0b5633ba2b904511ad514346f4d23d4b",
    "name": "serviceInstanceName1",
    "description": "description",
    "endpoint_id": "b935d0ef-f4eb-4b95-aff1-9d33ae9f57b6",
    "status": "",
    "create_time": "2023-05-30T12:24:30.401Z",
    "update_time": "2023-05-30T12:24:30.401Z",
    "duration": 0,
    "create_user": {
      "domain_name": "string",
      "user_name": "string"
    },
    "type": "LLM_MODEL",
    "error_code": "lakehouse.00010003",
    "error_msg": "failed",
    "solution": "failed",
    "config": {
      "model_instance_config": {
        "resource": {
          "min": 1,
```

```
    "max" : 1000,
    "spec_code" : "string"
  }
},
"content_moderation_switch" : "ON",
"content_output_moderation_switch" : "OFF"
}]
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	查询service instance列表的响应体。
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.12.2 查看部署的服务实例详情

功能介绍

查看部署后的服务实例的详情。返回服务实例的详情信息。此接口为同步接口，无配套使用接口和特殊场景。

URI

GET /v1/workspaces/{workspace_id}/services/instances/{instance_id}

表 4-453 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。
instance_id	是	String	参数解释： 实例的ID。获取方法，请参见 获取推理实例ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-454 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户token。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

响应参数

状态码：200

service部署信息。

状态码：400

表 4-455 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：401

表 4-456 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：404

表 4-457 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：408

表 4-458 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：500

表 4-459 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

请求示例

```
GET https://{endpoint}/v1/workspaces/{workspace_id}/services/instances/{instance_id}
```

响应示例

状态码：200

service部署信息。

```
{
  "visibility": "PRIVATE",
  "source": {
    "id": "0b5633ba2b904511ad514346f4d23d4b",
    "version_id": "0b5633ba2b904511ad514346f4d23d4b"
  },
  "id": "0b5633ba2b904511ad514346f4d23d4b",
  "name": "serviceInstanceName1",
  "description": "description",
  "endpoint_id": "b935d0ef-f4eb-4b95-aff1-9d33ae9f57b6",
  "status": "",
  "create_time": "2023-05-30T12:24:30.401Z",
  "update_time": "2023-05-30T12:24:30.401Z",
  "duration": 0,
  "create_user": {
    "domain_name": "string",
    "user_name": "string"
  },
  "type": "LLM_MODEL",
  "error_code": "lakehouse.00010003",
  "error_msg": "failed",
  "solution": "failed",
  "config": {
    "model_instance_config": {
      "resource": {
        "min": 1,
        "max": 1000,
        "spec_code": "string"
      }
    }
  },
  "content_moderation_switch": "ON",
  "content_output_moderation_switch": "ON",
  "update_user": {
    "domain_name": "string",
    "user_name": "string"
  },
  "url": {
    "name": "string",
    "address": "string",
    "type": "PRIVATE"
  }
}
```

状态码：400

BadRequest

```
{
  "error_code": "common.01000001",
  "error_msg": "failed to read http request, please check your input, code: 400, reason: Type mismatch., cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code": "APIG.1002",
  "error_msg": "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error": {
    "code": "403",
    "message": "X-Auth-Token is invalid in the request",
    "title": "Forbidden"
  },
  "error_code": 403,
  "error_msg": "X-Auth-Token is invalid in the request",
  "title": "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code": "common.01000001",
  "error_msg": "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code": "common.00000408",
  "error_msg": "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code": "common.00000500",
  "error_msg": "internal error"
}
```

状态码

状态码	描述
200	service部署信息。
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.12.3 部署服务

功能介绍

部署一个服务实例。部署服务实例接口支持输入名称、描述、端点ID、实例规格、最小值和最大值等参数，返回实例ID。此过程包含资源校验：最大值不小于最小值，且不超过端点最大资源数；同一端点下所有服务总资源数亦需符合限制。服务实例将依据负载自动调整数量，确保在最小与最大值范围内运行。

URI

POST /v1/workspaces/{workspace_id}/services/instances

表 4-460 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-461 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户token。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

表 4-462 请求 Body 参数

参数	是否必选	参数类型	描述
source	是	SourceRef object	参数解释： 来源信息，Service信息或者模型信息。 约束限制： 不涉及。

参数	是否必选	参数类型	描述
name	是	String	参数解释： 一个服务实例的名称。 约束限制： 不涉及。 取值范围： 长度为[1,64]的中文、字母、数字、下划线(_)、中划线(-)、点、空格的组合。 默认取值： 不涉及。
description	否	String	参数解释： 描述信息。 约束限制： 不涉及。 取值范围： 长度为[0,1024]的除了“!”、“<”、“>”、“=”、“&”、“”、“,”之外的任意字符的组合。 默认取值： 不涉及。
endpoint_id	是	String	参数解释： 端点ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。
config	否	ServiceInstanceConfig object	参数解释： 启动Service实例时输入的配置，在DeployService中使用。 约束限制： 不涉及。

表 4-463 SourceRef

参数	是否必选	参数类型	描述
id	是	String	参数解释： 来源ID。有不同的类型，对于推理服务实例，则为模型的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字母、数字、中划线(-)的组合。 默认取值： 不涉及。

参数	是否必选	参数类型	描述
version_id	否	String	参数解释： 版本ID。有不同的类型，对于推理服务实例，则为模型版本ID 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字母、数字、中划线(-)的组合。 默认取值： 不涉及。

表 4-464 ServiceInstanceConfig

参数	是否必选	参数类型	描述
model_instance_config	否	ModelServiceInstanceConfig object	参数解释： 模型服务实例的配置，在DeployService中使用。 约束限制： 不涉及。

表 4-465 ModelServiceInstanceConfig

参数	是否必选	参数类型	描述
resource	是	ResourceDemand object	参数解释： 资源需求量配置。 约束限制： 不涉及。

表 4-466 ResourceDemand

参数	是否必选	参数类型	描述
min	是	Integer	参数解释： 最小数。 约束限制： 不涉及。 取值范围： [1,1000]。 默认取值： 不涉及。
max	是	Integer	参数解释： 最大数。 约束限制： 不涉及。 取值范围： [1,1000]。 默认取值： 不涉及。

参数	是否必选	参数类型	描述
spec_code	是	String	参数解释： 资源规格，从规格列表查询获取。 约束限制： 不涉及。 取值范围： 长度为[1,64]。 默认取值： 不涉及。

响应参数

状态码：200

表 4-467 响应 Body 参数

参数	参数类型	描述
id	InstanceId object	参数解释： 实例ID。获取方法，请参见 获取推理实例ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。

表 4-468 InstanceId

参数	参数类型	描述
-	String	参数解释： 实例ID。获取方法，请参见 获取推理实例ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。

状态码：400

表 4-469 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：401

表 4-470 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：404

表 4-471 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：408

表 4-472 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：500

表 4-473 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

请求示例

部署一个服务实例，具体部署参数如下示例所示。

```
POST https://{endpoint}/v1/workspaces/{workspace_id}/services/instances
{
  "source": {
    "id": "0b5633ba2b904511ad514346f4d23d4b",
    "version_id": "0b5633ba2b904511ad514346f4d23d4b"
  },
  "name": "serviceInstanceName1",
  "description": "description",
  "endpoint_id": "b935d0ef-f4eb-4b95-aff1-9d33ae9f57b6",
  "config": {
    "model_instance_config": {
      "resource": {
        "min": 1,
        "max": 1000,
        "spec_code": "string"
      }
    }
  }
}
```

响应示例

状态码：400

BadRequest

```
{
  "error_code": "common.01000001",
  "error_msg": "failed to read http request, please check your input, code: 400, reason: Type mismatch., cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	OK
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out

状态码	描述
500	InternalServerError

错误码

请参见[错误码](#)。

4.12.4 更新已部署的服务实例

功能介绍

更新已部署的服务实例。用户可以通过该接口更新已部署的服务实例，包括名称和描述等。此接口为异步接口，配套使用接口ShowServiceInstanceDetail，ListServicesInstances来查询服务实例具体情况。目前暂不支持更新实例最小数和模型版本。

URI

PUT /v1/workspaces/{workspace_id}/services/instances/{instance_id}

表 4-474 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。
instance_id	是	String	参数解释： 实例的ID。获取方法，请参见 获取推理实例ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-475 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户token。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

表 4-476 请求 Body 参数

参数	是否必选	参数类型	描述
source	否	SourceRef object	参数解释：来源信息，Service 信息或者模型信息。 约束限制：不涉及。
name	否	String	参数解释：一个服务实例的名称。 约束限制：不涉及。 取值范围：长度为[1,64]的中文、字母、数字、下划线(_)、中划线(-)、点、空格的组合。 默认取值：不涉及。
description	否	String	参数解释：描述信息。 约束限制：不涉及。 取值范围：长度为[0,1024]的除了“!”、“<”、“>”、“=”、“&”、“”、“,”之外的任意字符的组合。 默认取值：不涉及。
config	否	UpdateServiceInstanceConfig object	参数解释：更新服务实例时输入的配置。 约束限制：不涉及。

表 4-477 SourceRef

参数	是否必选	参数类型	描述
id	是	String	参数解释： 来源ID。有不同的类型，对于推理服务实例，则为模型的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字母、数字、中划线(-)的组合。 默认取值： 不涉及。
version_id	否	String	参数解释： 版本ID。有不同的类型，对于推理服务实例，则为模型版本ID 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字母、数字、中划线(-)的组合。 默认取值： 不涉及。

表 4-478 UpdateServiceInstanceConfig

参数	是否必选	参数类型	描述
model_instance_config	否	UpdateModelServiceInstanceConfig object	参数解释： 更新模型服务实例的配置。 约束限制： 不涉及。

表 4-479 UpdateModelServiceInstanceConfig

参数	是否必选	参数类型	描述
resource	是	BaseDemand object	参数解释： 资源需求量配置。 约束限制： 不涉及。

表 4-480 BaseDemand

参数	是否必选	参数类型	描述
min	是	Integer	参数解释：最小数。 约束限制：不涉及。 取值范围：[1,1000]。 默认取值：不涉及。
max	是	Integer	参数解释：最大数。 约束限制：不涉及。 取值范围：[1,1000]。 默认取值：不涉及。

响应参数

状态码：202

表 4-481 响应 Body 参数

参数	参数类型	描述
id	InstanceId object	参数解释：实例ID。获取方法，请参见 获取推理实例ID 。 约束限制：不涉及。 取值范围：长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值：不涉及。

表 4-482 InstanceId

参数	参数类型	描述
-	String	参数解释：实例ID。获取方法，请参见 获取推理实例ID 。 约束限制：不涉及。 取值范围：长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值：不涉及。

状态码：400

表 4-483 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：401

表 4-484 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：404

表 4-485 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：408

表 4-486 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：500

表 4-487 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

请求示例

更新已部署的服务实例，具体更新参数如下示例所示。

```
PUT https://{endpoint}/v1/workspaces/{workspace_id}/services/instances/{instance_id}

{
  "source": {
    "id": "0b5633ba2b904511ad514346f4d23d4b",
    "version_id": "0b5633ba2b904511ad514346f4d23d4b"
  },
  "name": "serviceInstanceName1",
  "description": "description",
  "config": {
    "model_instance_config": {
      "resource": {
        "min": 1,
        "max": 1000
      }
    }
  }
}
```

响应示例

状态码：202

更新推理模型的响应。

```
{
  "id": "b935d0ef-f4eb-4b95-aff1-9d33ae9f57b6"
}
```

状态码：400

BadRequest

```
{
  "error_code": "common.01000001",
}
```

```
"error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,  
cause: TypeMismatchException"  
}
```

状态码：401

Unauthorized

```
{  
  "error_code" : "APIG.1002",  
  "error_msg" : "Incorrect token or token resolution failed"  
}
```

状态码：403

Forbidden

```
{  
  "error" : {  
    "code" : "403",  
    "message" : "X-Auth-Token is invalid in the request",  
    "title" : "Forbidden"  
  },  
  "error_code" : 403,  
  "error_msg" : "X-Auth-Token is invalid in the request",  
  "title" : "Forbidden"  
}
```

状态码：404

NotFound

```
{  
  "error_code" : "common.01000001",  
  "error_msg" : "response status exception, code: 404"  
}
```

状态码：408

Request Time-out

```
{  
  "error_code" : "common.00000408",  
  "error_msg" : "timeout exception occurred"  
}
```

状态码：500

InternalServerError

```
{  
  "error_code" : "common.00000500",  
  "error_msg" : "internal error"  
}
```

状态码

状态码	描述
202	更新推理模型的响应。
400	BadRequest
401	Unauthorized
403	Forbidden

状态码	描述
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.12.5 删除服务实例

功能介绍

删除服务实例，释放该实例的资源。用户可以通过该接口删除服务实例，释放该实例的资源。此接口为异步接口，配套使用接口ShowServiceInstanceDetail，ListServicesInstances来查询服务实例具体情况，无特殊场景。

URI

DELETE /v1/workspaces/{workspace_id}/services/instances/{instance_id}

表 4-488 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。
instance_id	是	String	参数解释： 实例的ID。获取方法，请参见 获取推理实例ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-489 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户token。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

响应参数

状态码：202

表 4-490 响应 Body 参数

参数	参数类型	描述
-	AnyType	删除Service的响应。

状态码：400

表 4-491 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：401

表 4-492 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：404

表 4-493 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：408

表 4-494 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：500

表 4-495 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

请求示例

GET https://{endpoint}/v1/workspaces/{workspace_id}/services/instances/{instance_id}

响应示例

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
202	删除Service的响应。
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.13 推理服务接口

4.13.1 发起调用请求

功能介绍

调用已部署的大模型推理实例，发起推理请求。此接口为同步接口，无配套使用接口。该接口会有内容审核，对于不符合要求的内容会进行屏蔽，用户可以选择是否关闭。

URI

POST /v1/workspaces/{workspace_id}/services/instances/{instance_id}/invocations

表 4-496 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。

参数	是否必选	参数类型	描述
instance_id	是	String	参数解释： 实例的ID。获取方法，请参见 获取推理实例ID 。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-497 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户token。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

表 4-498 请求 Body 参数

参数	是否必选	参数类型	描述
messages	否	Array of ChatMessage objects	参数解释： 消息。 约束限制： [1,100000]。
max_tokens	否	Integer	参数解释： 要在聊天完成中生成的最大token数。输入token和生成token的总长度受模型的上下文长度限制。输入0时模型按照默认值4096处理。R1模型取值范围为[0, 32k]， V3模型取值范围为[0, 16k]。不可与max_completion_tokens字段同时设置，会直接报错。 约束限制： 输入token和生成token的总长度受模型的上下文长度限制。 取值范围： 不涉及。 默认取值： 不涉及。

参数	是否必选	参数类型	描述
temperature	否	Double	参数解释： Temperature是用于调整随机程度的数字。介于0和2之间。较高的值（如0.8）将使输出更随机，而较低的值（如0.2）将使输出更集中和确定性。 约束限制： 不涉及。 取值范围： [0,2]。 默认取值： 1。
logit_bias	否	Object	参数解释： 接受一个map值，其中每个键为词表中的token ID（使用tokenization接口获取），为整数，每个值为该token的偏差值，为浮点数。调整指定token在模型输出内容中出现的概率，使模型生成的内容更加符合特定的偏好。该参数暂不支持。 约束限制： 不涉及。
top_p	否	Double	参数解释： 核心采样，用于控制AI模型根据累积概率考虑的标记范围。当取值为0时模型仅考虑对数概率最大的一个token。 约束限制： 不涉及。 取值范围： [0,1]。 默认取值： 1。
stream	否	Boolean	参数解释： 是否支持流式返回。如果支持，则消息按行返回（交互式效果）。如果不支持，则消息一次性全部返回。 约束限制： 不涉及。 取值范围： true, false。 默认取值： 不涉及。
frequency_penalty	否	Double	参数解释： 频率惩罚，控制文本中词汇的重复度，避免生成文本中某些词汇或短语出现过于频繁。正值会根据它们在文本中的现有频率惩罚新令牌，从而降低模型逐字重复同一行的可能性。 约束限制： 不涉及 取值范围： [-2.0,2.0]。 默认取值： 不涉及。

参数	是否必选	参数类型	描述
presence_penalty	否	Double	参数解释： 存在惩罚，控制文本中话题的重复度，避免在对话或文本中反复讨论相同的主题或观点。正值会根据到目前为止它们是否出现在文本中来惩罚新令牌，从而增加模型谈论新主题的可能性。 约束限制： 不涉及 取值范围： [-2.0,2.0]。 默认取值： 不涉及。
n	否	Integer	参数解释： 要为每个输入消息生成多少个聊天完成选项。请注意，您将根据所有选项中生成的token数收取费用。将n保持为1，以最小化成本。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

表 4-499 ChatMessage

参数	是否必选	参数类型	描述
role	是	String	参数解释： 角色。 约束限制： 不涉及。 取值范围： 长度为[1,64]的除了“!”、“<”、“>”、“=”、“&”、“”、“,”之外的任意字符的组合。 默认取值： 不涉及。
content	否	Object	参数解释： 消息的内容，可以是str或object[]。 约束限制： 不涉及。
name	否	String	提供模型信息，用于区分同一角色的参与者。
tool_calls	否	Array of MessageTool Call objects	模型生成的工具调用，content与tool_calls字段二者至少有一个为非空。
tool_call_id	否	String	模型生成的工具调用id

表 4-500 MessageToolCall

参数	是否必选	参数类型	描述
id	是	String	当前工具调用ID。
type	是	String	tool的类型。目前仅支持function。
function	是	Function object	模型调用的function。

表 4-501 Function

参数	是否必选	参数类型	描述
name	是	String	模型调用的function名。
arguments	是	String	要调用的function的参数，由模型生成，格式为JSON。请注意，模型并不总是生成有效的JSON，并且可能会臆造出你函数模式中未定义的参数。在调用函数之前，请在代码中验证这些参数。

响应参数

状态码：200

创建ChatCompletions的响应体。

状态码：400

表 4-502 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：401

表 4-503 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：404

表 4-504 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：408

表 4-505 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：500

表 4-506 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释： 解决方案描述。 约束限制： 不涉及。 取值范围： 长度[2,4096]。 默认取值： 不涉及。

请求示例

调用已部署的大模型推理实例，发起推理请求。具体请求参数如下示例所示。

```
POST https://{endpoint}/v1/workspaces/{workspace_id}/services/instances/{instance_id}/invocations
{
  "messages": [{
    "role": "user",
    "content": "请总结2023年LLM的发展"
  }]
}
```

响应示例

状态码：200

创建ChatCompletions的响应体。

```
{
  "route_id": "ac8111bf-3601-4905-8ddd-b41d3e636a4e"
}
```

状态码：400

BadRequest

```
{
  "error_code": "common.01000001",
  "error_msg": "failed to read http request, please check your input, code: 400, reason: Type mismatch., cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code": "APIG.1002",
  "error_msg": "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error": {
    "code": "403",
    "message": "X-Auth-Token is invalid in the request",
    "title": "Forbidden"
  },
  "error_code": 403,
  "error_msg": "X-Auth-Token is invalid in the request",
}
```

```
"title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	创建ChatCompletions的响应体。
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.14 Ray Session 接口

4.14.1 获取 Session 信息

功能介绍

获取Session信息。ray和ray serve场景下查看dashboard时使用；输入ray或ray serve集群的路由ID；输出session ID。此接口为同步接口，无配套使用接口和特殊场景。

URI

POST /v1/sessions

表 4-507 Query 参数

参数	是否必选	参数类型	描述
route_id	是	String	参数解释： 必填路由ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-508 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户token。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

响应参数

状态码：200

表 4-509 响应 Body 参数

参数	参数类型	描述
session_id	String	参数解释： Session的ID。 约束限制： 不涉及。 取值范围： 长度为[1,32]的英文字符、数字和中划线(-)等特殊符号的组合。 默认取值： 不涉及。

参数	参数类型	描述
status	String	参数解释：session的状态。 约束限制：不涉及。 取值范围：[0,64]。 默认取值：不涉及。

状态码：400

表 4-510 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：401

表 4-511 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：404

表 4-512 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：408

表 4-513 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：500

表 4-514 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

请求示例

获取session信息。无请求体。

POST https://{endpoint}/v1/sessions

响应示例

状态码：200

获取Session信息的响应。

```
{
  "session_id" : "ac8111bf-3601-4905-8ddd-b41d3e636a4e",
  "status" : "RUNNING"
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
}
```

```
"error_msg": "failed to read http request, please check your input, code: 400, reason: Type mismatch.,  
cause: TypeMismatchException"  
}
```

状态码：401

Unauthorized

```
{  
  "error_code": "APIG.1002",  
  "error_msg": "Incorrect token or token resolution failed"  
}
```

状态码：403

Forbidden

```
{  
  "error": {  
    "code": "403",  
    "message": "X-Auth-Token is invalid in the request",  
    "title": "Forbidden"  
  },  
  "error_code": 403,  
  "error_msg": "X-Auth-Token is invalid in the request",  
  "title": "Forbidden"  
}
```

状态码：404

NotFound

```
{  
  "error_code": "common.01000001",  
  "error_msg": "response status exception, code: 404"  
}
```

状态码：408

Request Time-out

```
{  
  "error_code": "common.00000408",  
  "error_msg": "timeout exception occurred"  
}
```

状态码：500

InternalServerError

```
{  
  "error_code": "common.00000500",  
  "error_msg": "internal error"  
}
```

状态码

状态码	描述
200	获取Session信息的响应。
400	BadRequest
401	Unauthorized
403	Forbidden

状态码	描述
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.15 SQL 业务接口

4.15.1 创建 SQL 的 Session

功能介绍

创建SQL Session。用户通过此接口在某SQL端点创建SQL Session，通过输入端点id、LakeFormation配置，返回Session信息。此接口为同步接口，无配套使用接口和特殊场景。

URI

POST /v1/workspaces/{workspace_id}/sessions

表 4-515 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-516 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户token。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

表 4-517 请求 Body 参数

参数	是否必选	参数类型	描述
endpoint_id	是	String	参数解释：端点ID。 约束限制：不涉及。 取值范围：长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值：不涉及。
lake_formation_config	是	LakeFormationInfo object	参数解释：LakeFormation相关信息。 约束限制：不涉及。
wait_timeout	否	Integer	参数解释：创建session排队等待时间，单位秒。 约束限制：不涉及。 取值范围：[300,7200]。 默认取值：不涉及。

表 4-518 LakeFormationInfo

参数	是否必选	参数类型	描述
instance_id	是	String	参数解释：lakeformation实例ID。 约束限制：不涉及。 取值范围：长度为[1,64]的英文字符、数字和中划线(-)的组合。 默认取值：不涉及。

参数	是否必选	参数类型	描述
catalog	是	String	参数解释： lakeformation的catalog。 约束限制： 不涉及。 取值范围： 长度为[1,255]的英文字符、数字、中划线(-)和下划线(_)的组合。 默认取值： 不涉及。

响应参数

状态码：200

表 4-519 响应 Body 参数

参数	参数类型	描述
session_id	String	参数解释： Session的ID。 约束限制： 不涉及。 取值范围： 长度为[1,32]的英文字符、数字和中划线(-)等特殊符号的组合。 默认取值： 不涉及。
status	String	参数解释： session的状态。 约束限制： 不涉及。 取值范围： [0,64]。 默认取值： 不涉及。

状态码：400

表 4-520 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释： 错误码。 约束限制： 不涉及。 取值范围： 长度[8,36]。 默认取值： 不涉及。
error_msg	String	参数解释： 错误描述。 约束限制： 不涉及。 取值范围： 长度[2,4096]。 默认取值： 不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：401

表 4-521 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：404

表 4-522 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：408

表 4-523 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：500

表 4-524 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释： 解决方案描述。 约束限制： 不涉及。 取值范围： 长度[2,4096]。 默认取值： 不涉及。

请求示例

创建SQL Session。用户通过此接口在某SQL端点创建SQL Session，通过输入端点id、LakeFormation配置，返回Session信息。此接口为同步接口，无配套使用接口和特殊场景。

```
POST https://endpoint/v1/workspaces/{workspace_id}/sessions
{
  "endpoint_id": "b935d0ef-f4eb-4b95-aff1-9d33ae9f57b6",
  "lake_formation_config": {
    "instance_id": "0b5633ba2b904511ad514346f4d23d4b",
    "catalog": "string"
  }
}
```

响应示例

状态码：200

获取Session信息的响应。

```
{
  "session_id": "ac8111bf-3601-4905-8ddd-b41d3e636a4e",
  "status": "RUNNING"
}
```

状态码：400

BadRequest

```
{
  "error_code": "common.01000001",
  "error_msg": "failed to read http request, please check your input, code: 400, reason: Type mismatch., cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code": "APIG.1002",
  "error_msg": "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error": {
    "code": "403",
    "message": "X-Auth-Token is invalid in the request",
  }
}
```

```
{
  "title" : "Forbidden"
},
"error_code" : 403,
"error_msg" : "X-Auth-Token is invalid in the request",
"title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	获取Session信息的响应。
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.15.2 获取 SQL Session 列表

功能介绍

查询SQL Session列表

URI

GET /v1/workspaces/{workspace_id}/sessions

表 4-525 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。

表 4-526 Query 参数

参数	是否必选	参数类型	描述
endpoint_id	否	String	参数解释： 通过ID检索Endpoint的参数。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。
status	否	String	参数解释： 状态过滤，支持一种状态查询，默认查询所有。 约束限制： 不涉及。 取值范围： 可选值有：RUNNING、CLOSED、WAITING、CREATING、FAIL 默认取值： 不涉及。
session_id	否	String	参数解释： SessionId的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。
limit	否	Integer	参数解释： 指定每一页返回的最大条目数。 约束限制： 不涉及。 取值范围： [1,100]。 默认取值： 10。

参数	是否必选	参数类型	描述
offset	否	Integer	参数解释： 偏移量，表示从此偏移量开始查询。 约束限制： 不涉及。 取值范围： [0,1000000000]。 默认取值： 0。

请求参数

表 4-527 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户token。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

响应参数

状态码：200

表 4-528 响应 Body 参数

参数	参数类型	描述
sessions	Array of SqlSessionInfo objects	参数解释： Session列表。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
total	Integer	总数

表 4-529 SqlSessionInfo

参数	参数类型	描述
session_id	String	参数解释： Session的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)等特殊字符的组合。 默认取值： 不涉及。

参数	参数类型	描述
status	String	参数解释： • RUNNING-运行中 • CLOSED-已关闭 • WAITING-等待中 • CREATING-创建中 • FAIL-失败 约束限制： 不涉及。 取值范围： RUNNING、CLOSED、WAITING、CREATING、FAIL 默认取值： 不涉及。
create_time	String	参数解释： session创建时间。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。
endpoint_info	endpoint_info object	参数解释： 创建Session的端点信息。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

表 4-530 endpoint_info

参数	参数类型	描述
id	String	参数解释： 端点ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。
name	String	参数解释： Endpoint名称。 约束限制： 只能包含中文、字母、数字、下划线、中划线、点、空格。 取值范围： 不涉及。 默认取值： 不涉及。

状态码：400

表 4-531 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：401

表 4-532 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：404

表 4-533 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：408

表 4-534 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：500

表 4-535 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

请求示例

查询Session列表信息，返回值为Session状态、创建时间、端点等信息

```
GET https://endpoint/v1/workspaces/{workspace_id}/sessions?endpoint_id=
{ }
```

响应示例

状态码：200

Session列表响应体

```
{
  "sessions": [ {
    "session_id": "284888e2-c36f-4bec-9b24-e5bf0b386ab0",
    "status": "RUNINNG",
    "create_time": "2025-07-22T08:42:39.000+00:00",
    "endpoint_info": {
      "id": "0470d220-aaaa-410e-a417-0d69a0215683",
      "name": "test2"
    }
  }, {
    "session_id": "4e4b2770-702f-48b7-9737-db107cf63e24",
    "status": "CLOSED",
    "create_time": "2025-07-22T06:13:47.000+00:00",
    "endpoint_info": {
      "id": "0470d220-aaaa-410e-a417-0d69a0215683",
      "name": "test2"
    }
  } ],
  "total": 15
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error": {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	Session列表响应体
400	BadRequest
401	Unauthorized

状态码	描述
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.15.3 查询 Session 信息

功能介绍

查询SQL Session信息。用户可通过此接口查询SQL Session信息，输入为Session id，返回操作结果。此接口为同步接口，无配套使用接口和特殊场景。

URI

GET /v1/workspaces/{workspace_id}/sessions/{session_id}

表 4-536 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。
session_id	是	String	参数解释： SessionId的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-537 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户token。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

响应参数

状态码：200

表 4-538 响应 Body 参数

参数	参数类型	描述
session_id	String	参数解释：Session的ID。 约束限制：不涉及。 取值范围：长度为[1,32]的英文字符、数字和中划线(-)等特殊符号的组合。 默认取值：不涉及。
status	String	参数解释：session的状态。 约束限制：不涉及。 取值范围：[0,64]。 默认取值：不涉及。

状态码：400

表 4-539 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：401

表 4-540 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：404

表 4-541 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：408

表 4-542 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：500

表 4-543 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

参数	参数类型	描述
solution_msg	String	参数解释： 解决方案描述。 约束限制： 不涉及。 取值范围： 长度[2,4096]。 默认取值： 不涉及。

请求示例

查询SQL Session的信息。用户可通过该接口查询SQL语句执行结果，输入为端点id、SQL Session id，返回执行结果。此接口为同步接口，无配套使用接口和特殊场景。

```
GET https://endpoint/v1/workspaces/{workspace_id}/sessions/{session_id}
{ }
```

响应示例

状态码：200

获取Session信息的响应。

```
{
  "session_id" : "ac8111bf-3601-4905-8ddd-b41d3e636a4e",
  "status" : "RUNNING"
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch., cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	获取Session信息的响应。
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.15.4 关闭 Session

功能介绍

关闭SQL Session。用户可通过此接口关闭SQL Session，输入为Session id，返回操作结果。此接口为同步接口，无配套使用接口和特殊场景。

URI

DELETE /v1/workspaces/{workspace_id}/sessions/{session_id}

表 4-544 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释：工作空间的ID。 约束限制：不涉及。 取值范围：长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值：不涉及。
session_id	是	String	参数解释：SessionId的ID。 约束限制：不涉及。 取值范围：长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值：不涉及。

请求参数

表 4-545 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户token。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

响应参数

状态码：200

OK

状态码：400

表 4-546 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：401**表 4-547** 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：404**表 4-548** 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：408**表 4-549** 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：500**表 4-550** 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。

参数	参数类型	描述
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

请求示例

关闭SQL Session。用户可通过此接口删除SQL Session，输入为Session Id，返回操作结果。此接口为同步接口，无配套使用接口和特殊场景。

```
DELETE https://endpoint/v1/workspaces/{workspace_id}/sessions/{session_id}
{ }
```

响应示例

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	OK
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.15.5 执行 SQL

功能介绍

在指定的Session下执行SQL语句。用户可通过此接口执行SQL语句，输入为SQL语句、端点id、SQL Session id、绑定参数等信息，返回语句执行结果或语句id。此接口支持异步和同步两种执行模式，由入参is_sync参数决定。如果是异步执行，配套使用接口ShowSqlStatementResult查询语句结果。如果是同步执行，接口会等待一段时间接收结果，如果执行完毕会返回执行结果，如果没有执行完毕，会返回一个运行中状态，后续可以使用接口ShowSqlStatementResult查询语句结果。

URI

POST /v1/workspaces/{workspace_id}/sessions/statements

表 4-551 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。

请求参数

表 4-552 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户token。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

表 4-553 请求 Body 参数

参数	是否必选	参数类型	描述
statement	是	String	参数解释： SQL statement信息。 约束限制： 不涉及。 取值范围： [0,10000000]。 默认取值： 不涉及。
limit	否	Integer	参数解释： limit限制。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

参数	是否必选	参数类型	描述
session_id	是	String	参数解释： Session的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的英文字符、数字和中划线(-)等特殊字符的组合。 默认取值： 不涉及。
bindings	否	Array<Array<String>>	参数解释： 参数绑定列表。 约束限制： 不涉及。 取值范围： [0,10000]。 默认取值： 不涉及。
is_sync	否	Boolean	参数解释： 同步执行或异步执行。 约束限制： 不涉及。 取值范围： true：同步，false：异步。 默认取值： 不涉及。
is_maintain_statement	否	Boolean	参数解释： 是否来自运维通道。 约束限制： 不涉及。 取值范围： true：运维通道语句，false：不是运维通道语句。 默认取值： 不涉及。
resultset_fetch_mode	否	String	参数解释： 获取结果集的方式。 约束限制： 不涉及。 取值范围： DEFAULT：直接返回结果集；READ_OBS：返回结果集在桶上的路径，而不是直接返回结果集。 默认取值： DEFAULT。

响应参数

状态码：200

表 4-554 响应 Body 参数

参数	参数类型	描述
-	String	执行SQL的响应信息，String格式，可见响应示例。

状态码：400

表 4-555 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：401

表 4-556 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：404

表 4-557 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：408

表 4-558 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：500

表 4-559 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

请求示例

在指定的Session下执行SQL语句。用户可通过此接口执行SQL语句，输入为SQL语句、端点id、SQL Session id、绑定参数等信息，返回语句执行结果或语句id。此接口支持异步和同步两种执行模式，由入参is_sync参数决定。如果是异步执行，配套使用接口ShowSqlStatementResult查询语句结果。如果是同步执行，接口会等待一段时间接收结果，如果执行完毕会返回执行结果，如果没有执行完毕，会返回一个运行中状态，后续可以用使用接口ShowSqlStatementResult查询语句结果。

```
POST https://endpoint/v1/workspaces/{workspace_id}/sessions/statements
{
  "statement": "select * from aaa",
  "limit": 100,
  "session_id": "string",
  "bindings": [ [ "string" ] ],
  "is_maintain_statement": false,
  "resultset_fetch_mode": "DEFAULT",
  "is_sync": true
}
```

响应示例

状态码：200

OK

```
{
  "status": 1,
  "session_id": "6399466e-8908-4c9d-8399-48479c8b06f7",
  "statement_id": "49131d25-1824-4f29-860c-832bdd19dfd3"
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error": {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	OK
400	BadRequest
401	Unauthorized

状态码	描述
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.15.6 获取 Statement 结果、执行状态

功能介绍

查询SQL语句执行结果。用户可通过该接口查询SQL语句执行结果，输入为端点id、SQL Session id、语句id，返回执行结果。此接口为同步接口，无配套使用接口和特殊场景。

URI

GET /v1/workspaces/{workspace_id}/sessions/{session_id}/statements/{statement_id}

表 4-560 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。
session_id	是	String	参数解释： SessionId的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。
statement_id	是	String	参数解释： statement id。 约束限制： 不涉及。 取值范围： 长度为[1,64]的字母、数字、中划线(-)和下划线(_)的组合。 默认取值： 不涉及。

表 4-561 Query 参数

参数	是否必选	参数类型	描述
page_num	否	Integer	参数解释：查询页码。 约束限制：不涉及。 取值范围：[1,65535]。 默认取值：不涉及。
is_enable_obs_path	否	Boolean	参数解释：是否返回结果集的obs路径。 约束限制：不涉及。 取值范围：true：返回结果集的obs路径，false：返回结果集。 默认取值：不涉及。

请求参数

表 4-562 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户token。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

响应参数

状态码：200

表 4-563 响应 Body 参数

参数	参数类型	描述
-	String	SQL执行结果，String格式，可见响应示例。

状态码：400

表 4-564 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：401

表 4-565 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：404

表 4-566 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：408

表 4-567 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：500

表 4-568 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

请求示例

查询SQL语句执行结果。用户可通过该接口查询SQL语句执行结果，输入为端点id、SQL Session id、语句id，返回执行结果。此接口为同步接口，无配套使用接口和特殊场景。

```
GET https://endpoint/v1/workspaces/{workspace_id}/sessions/{session_id}/statements/{statement_id}
{ }
```

响应示例

状态码：200

OK

```
{
  "status" : 0,
  "session_id" : "6399466e-8908-4c9d-8399-48479c8b06f7",
  "statement_id" : "49131d25-1824-4f29-860c-832bdd19dfd3",
  "results" : [ {
    "status" : "PGRES_TUPLES_OK",
    "statement_id" : "49131d25-1824-4f29-860c-832bdd19dfd3",
    "num_rows" : 2,
    "row_count" : 2,
    "page_count" : 1,
    "page_no" : 1,
    "err_code" : "0",
    "sql_state" : "0",
    "message" : "",
    "result_set" : {
      "columns" : [ {
        "name" : "a",
        "table_id" : 2147483647,
        "column_id" : 1,
        "format" : 0,
        "type" : 23,
        "size" : 4,
```

```
    "type_mod" : -1
  }, {
    "name" : "b",
    "table_id" : 2147483647,
    "column_id" : 2,
    "format" : 0,
    "type" : 23,
    "size" : 4,
    "type_mod" : -1
  } ],
  "rows" : [ [ "1", "2" ], [ "1", "2" ] ]
}
}]
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch., cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	OK
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.15.7 取消 Statement 执行

功能介绍

取消SQL语句执行。用户可通过该接口取消SQL语句执行，输入为端点id、SQL Session id、语句id，返回操作结果。此接口为同步接口，无配套使用接口和特殊场景。

URI

POST /v1/workspaces/{workspace_id}/sessions/{session_id}/statements/{statement_id}/cancel

表 4-569 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。

参数	是否必选	参数类型	描述
session_id	是	String	参数解释： SessionId的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。
statement_id	是	String	参数解释： statement id。 约束限制： 不涉及。 取值范围： 长度为[1,64]的字母、数字、中划线(-)和下划线(_)的组合。 默认取值： 不涉及。

请求参数

表 4-570 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户token。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

响应参数

状态码：200

表 4-571 响应 Body 参数

参数	参数类型	描述
-	AnyType	空响应。

状态码：400

表 4-572 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：401

表 4-573 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：404

表 4-574 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：408

表 4-575 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：500

表 4-576 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

请求示例

取消SQL语句执行。用户可通过该接口取消SQL语句执行，输入为端点id、SQL Session id、语句id，返回操作结果。此接口为同步接口，无配套使用接口和特殊场景。

```
POST https://endpoint/v1/workspaces/{workspace_id}/sessions/{session_id}/cancel
{ }
```

响应示例

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch., cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
```

```
"code" : "403",  
"message" : "X-Auth-Token is invalid in the request",  
"title" : "Forbidden"  
},  
"error_code" : 403,  
"error_msg" : "X-Auth-Token is invalid in the request",  
"title" : "Forbidden"  
}
```

状态码：404

NotFound

```
{  
  "error_code" : "common.01000001",  
  "error_msg" : "response status exception, code: 404"  
}
```

状态码：408

Request Time-out

```
{  
  "error_code" : "common.00000408",  
  "error_msg" : "timeout exception occurred"  
}
```

状态码：500

InternalServerError

```
{  
  "error_code" : "common.00000500",  
  "error_msg" : "internal error"  
}
```

状态码

状态码	描述
200	空响应。
400	BadRequest
401	Unauthorized
403	Forbidden
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.15.8 获取 SQL 执行记录

功能介绍

查询SQL执行记录

URI

GET /v1/workspaces/{workspace_id}/statement-records

表 4-577 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。

表 4-578 Query 参数

参数	是否必选	参数类型	描述
endpoint_id	否	String	参数解释： 通过ID检索Endpoint的参数。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。
statement_id	否	String	参数解释： statement id。 约束限制： 不涉及。 取值范围： 长度为[1,64]的字母、数字、中划线(-)和下划线(_)的组合。 默认取值： 不涉及。
limit	否	Integer	参数解释： 指定每一页返回的最大条目数。 约束限制： 不涉及。 取值范围： [1,100]。 默认取值： 10。

参数	是否必选	参数类型	描述
offset	否	Integer	参数解释：偏移量，表示从此偏移量开始查询。 约束限制：不涉及。 取值范围：[0,1000000000]。 默认取值：0。
source	否	Integer	参数解释：执行sql来源，0 API，1 Console。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
start_time	否	Integer	参数解释：查询开始时间。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
end_time	否	Integer	参数解释：查询完成时间。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

请求参数

表 4-579 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释：租户token。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

响应参数

状态码：200

表 4-580 响应 Body 参数

参数	参数类型	描述
total	Integer	参数解释：符合条件的sql记录总数。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
records	Array of StatementRecord objects	sql历史记录。

表 4-581 StatementRecord

参数	参数类型	描述
id	String	Sql执行StatementID。
workspace_id	String	参数解释：工作空间ID。 约束限制：不涉及。 取值范围：长度为[1,36]的英文字符、数字和中划线(-)的组合。 默认取值：不涉及。
endpoint_id	String	参数解释：端点ID。 约束限制：不涉及。 取值范围：长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值：不涉及。
database_name	String	参数解释：数据库名称。 约束限制：不涉及。 取值范围：长度为[1,256]的字母、数字、下划线(_)的组合。 默认取值：不涉及。
statement_content	String	参数解释：sql脚本内容。 约束限制：不涉及。 取值范围：长度为[1,1024]的中文、字母、数字、中划线(-)、下划线(_)的组合。 默认取值：不涉及。
dpu_duration	String	参数解释：dpu时。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

参数	参数类型	描述
duration	String	参数解释：执行时长。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
dpu_cost	String	参数解释：费用。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
status	String	参数解释：Sql执行状态。 - CANCELED取消 - FAILED失败 - SUCCESSFUL成功 - RUNNING运行中 - SUBMITTED提交中 - ERROR错误 约束限制：不涉及。 取值范围：CANCELED FAILED SUCCESSFUL RUNNING SUBMITTED ERROR。 默认取值：不涉及。
user_name	String	参数解释：用户名称。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
create_time	String	参数解释：创建时间。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
start_time	String	参数解释：sql执行开始时间。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
finish_time	String	参数解释：sql执行结束时间。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

状态码：400

表 4-582 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：401

表 4-583 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：404

表 4-584 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：408

表 4-585 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：500

表 4-586 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

请求示例

查询sql执行历史记录，返回值为sql执行的sql，开始时间，结束时间等。

```
GET https://endpoint/v1/workspaces/{workspace_id}/statement-records?endpoint_id=
{ }
```

响应示例

状态码：200

Sql执行历史记录响应体。

```
{
  "total": 1,
  "records": [ {
    "id": "0b5633ba2b904511ad514346f4dxxxxx",
    "workspace_id": "0b5633ba2b904511ad514346f4d23d4b",
    "endpoint_id": "0b5633ba2b904511ad514346f4d23d4b",
    "database_name": "db1",
    "statement_content": "select 1",
    "dpu_duration": 2.23,
    "dpu_cost": 10.11,
    "duration": 332,
    "status": "SUCCESSFUL",
    "create_time": "2023-05-30T12:24:30.401Z",
    "start_time": "2023-05-30T12:24:30.401Z",
    "finish_time": "2023-05-30T12:24:30.401Z"
  } ]
}
```

状态码：400

BadRequest

```
{
  "error_code": "common.01000001",
}
```

```
"error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,  
cause: TypeMismatchException"  
}
```

状态码：401

Unauthorized

```
{  
  "error_code" : "APIG.1002",  
  "error_msg" : "Incorrect token or token resolution failed"  
}
```

状态码：403

Forbidden

```
{  
  "error" : {  
    "code" : "403",  
    "message" : "X-Auth-Token is invalid in the request",  
    "title" : "Forbidden"  
  },  
  "error_code" : 403,  
  "error_msg" : "X-Auth-Token is invalid in the request",  
  "title" : "Forbidden"  
}
```

状态码：404

NotFound

```
{  
  "error_code" : "common.01000001",  
  "error_msg" : "response status exception, code: 404"  
}
```

状态码：408

Request Time-out

```
{  
  "error_code" : "common.00000408",  
  "error_msg" : "timeout exception occurred"  
}
```

状态码：500

InternalServerError

```
{  
  "error_code" : "common.00000500",  
  "error_msg" : "internal error"  
}
```

状态码

状态码	描述
200	Sql执行历史记录响应体。
400	BadRequest
401	Unauthorized
403	Forbidden

状态码	描述
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

4.15.9 获取指定 Session 下的 SQL 执行记录

功能介绍

查询指定Session下的SQL执行记录

URI

GET /v1/workspaces/{workspace_id}/sessions/{session_id}/statements

表 4-587 路径参数

参数	是否必选	参数类型	描述
workspace_id	是	String	参数解释： 工作空间的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。
session_id	是	String	参数解释： SessionId的ID。 约束限制： 不涉及。 取值范围： 长度为[1,36]的字母、数字、中划线(-)的组合。 默认取值： 不涉及。

表 4-588 Query 参数

参数	是否必选	参数类型	描述
limit	否	Integer	参数解释： 指定每一页返回的最大条目数。 约束限制： 不涉及。 取值范围： [1,100]。 默认取值： 10。

参数	是否必选	参数类型	描述
offset	否	Integer	参数解释： 偏移量，表示从此偏移量开始查询。 约束限制： 不涉及。 取值范围： [0,1000000000]。 默认取值： 0。
statement_id	否	String	参数解释： statement id。 约束限制： 不涉及。 取值范围： 长度为[1,64]的字母、数字、中划线(-)和下划线(_)的组合。 默认取值： 不涉及。
status	否	String	参数解释： 状态过滤，支持一种状态查询，默认查询所有。 约束限制： 不涉及。 取值范围： 可选值有：CANCELED、FAILED、SUCCESSFUL、RUNNING、SUBMITTED、ERROR 默认取值： 不涉及。

请求参数

表 4-589 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	否	String	参数解释： 租户token。 约束限制： 不涉及。 取值范围： 不涉及。 默认取值： 不涉及。

响应参数

状态码：200

表 4-590 响应 Body 参数

参数	参数类型	描述
records	Array of records objects	参数解释：查询结果。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
total	Integer	参数解释：记录总数。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

表 4-591 records

参数	参数类型	描述
statement_id	String	参数解释：statement id。 约束限制：不涉及。 取值范围：长度为[1,64]的字母、数字、中划线(-)和下划线(_)的组合。 默认取值：不涉及。
statement_content	String	参数解释：查询语句内容。 约束限制：不涉及。 取值范围：长度为[0-2147483647]的字符串。 默认取值：不涉及。
create_time	String	参数解释：查询语句创建时间。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。
status	String	参数解释：查询状态。 约束限制：不涉及。 取值范围：CANCELED、FAILED、SUCCESSFUL、RUNNING、SUBMITTED、ERROR 默认取值：不涉及。
user_name	String	参数解释：用户名称。 约束限制：不涉及。 取值范围：不涉及。 默认取值：不涉及。

状态码：400

表 4-592 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：401

表 4-593 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：404

表 4-594 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：408

表 4-595 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

状态码：500

表 4-596 响应 Body 参数

参数	参数类型	描述
error_code	String	参数解释：错误码。 约束限制：不涉及。 取值范围：长度[8,36]。 默认取值：不涉及。
error_msg	String	参数解释：错误描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。
solution_msg	String	参数解释：解决方案描述。 约束限制：不涉及。 取值范围：长度[2,4096]。 默认取值：不涉及。

请求示例

查询sql执行历史记录，返回值为sql执行的sql，开始时间，状态等。

```
GET https://endpoint/v1/workspaces/{workspace_id}/sessions/{session_id}/statements
{ }
```

响应示例

状态码：200

获取Session下的SQL查询响应体

```
{
  "records" : [ {
    "statement_id" : "4272****02409",
    "statement_content" : "select 1",
    "create_time" : "2025-07-22T08:44:00.000+00:00",
    "status" : "FAILED",
    "user_name" : "example"
  }, {
    "statement_id" : "ba5e****78fd",
    "statement_content" : "select 1",
    "create_time" : "2025-07-22T08:43:58.000+00:00",
    "status" : "SUCCESSFUL",
    "user_name" : "example"
  } ],
  "total" : 2
}
```

状态码：400

BadRequest

```
{
  "error_code" : "common.01000001",
  "error_msg" : "failed to read http request, please check your input, code: 400, reason: Type mismatch.,
```

```
cause: TypeMismatchException"
}
```

状态码：401

Unauthorized

```
{
  "error_code" : "APIG.1002",
  "error_msg" : "Incorrect token or token resolution failed"
}
```

状态码：403

Forbidden

```
{
  "error" : {
    "code" : "403",
    "message" : "X-Auth-Token is invalid in the request",
    "title" : "Forbidden"
  },
  "error_code" : 403,
  "error_msg" : "X-Auth-Token is invalid in the request",
  "title" : "Forbidden"
}
```

状态码：404

NotFound

```
{
  "error_code" : "common.01000001",
  "error_msg" : "response status exception, code: 404"
}
```

状态码：408

Request Time-out

```
{
  "error_code" : "common.00000408",
  "error_msg" : "timeout exception occurred"
}
```

状态码：500

InternalServerError

```
{
  "error_code" : "common.00000500",
  "error_msg" : "internal error"
}
```

状态码

状态码	描述
200	获取Session下的SQL查询响应体
400	BadRequest
401	Unauthorized
403	Forbidden

状态码	描述
404	NotFound
408	Request Time-out
500	InternalServerError

错误码

请参见[错误码](#)。

5 应用示例

5.1 创建一个推理服务

概述

本文指导用户通过调用API的方式，介绍创建一个推理服务的完整流程。

本流程假设终端租户已经在Console界面完成DataArtsFabric的服务授权。API的调用方法请参见[如何调用API](#)。

前提条件

hostname：通过[地区和终端节点](#)查询服务的终端节点获取。

操作步骤

1. 调用创建工作空间接口，创建一个工作空间，记录接口返回工作空间的ID。

请求示例：

```
POST https://{hostname}/v1/workspaces
```

Body：

```
{
  "name": "apiworkspace",
  "description": "apie test workspace"
}
```

响应示例：

```
{
  "id": "e935d0ef-f4eb-4b95-aff1-9d33ae9f57a6",
  "name": "fabric",
  "description": "fabric",
  "create_time": "2023-05-30T12:24:30.401Z",
  "create_domain_name": "admin",
  "create_user_name": "user",
  "metastore_id": "2180518f-42b8-4947-b20b-adfc53981a25",
  "access_url": "https://test.fabric.com/",
  "enterprise_project_id": "01049549-82cd-4b2b-9733-ddb94350c125"
}
```

2. 调用[创建端点](#)创建端点接口，创建推理端点，记录接口返回的终端节点ID。

请求示例：

```
POST https://{hostname}/v1/workspaces/{workspace_id}/endpoints
```

workspace_id: 为步骤1中记录的工作空间ID。

Body:

```
{
  "name": "apie_test",
  "description": "apie test endpoint",
  "type": "inference",
  "reserved_resource": {
    "mu": {
      "spec_code": "mu.llama3.8b",
      "min": 0,
      "max": 1
    }
  }
}
```

响应示例:

```
{
  "visibility": "PRIVATE",
  "id": "0b5633ba2b904511ad514346f4d23d4b",
  "name": "endpoint1",
  "type": "inference",
  "status": "CREATING",
  "description": "description",
  "create_time": "2023-05-30T12:24:30.401Z",
  "update_time": "2023-05-30T12:24:30.401Z",
  "owner": {
    "domain_name": "string",
    "domain_id": "xxx",
    "user_name": "string",
    "user_id": "xxx"
  }
  "reserved_resource": {
    "mu": {
      "spec_code": "mu.llama3.8b",
      "min": 0,
      "max": 1
    }
  }
}
```

3. 调用[创建模型](#)创建模型接口，创建一个用户私有的模型，并记录接口返回模型ID。

请求示例:

POST https://{hostname}/v1/workspaces/{workspace_id}/models

workspace_id: 为步骤1中记录的工作空间ID。

Body:

```
{
  "name": "LLama3-8b",
  "description": "this is a apie test model",
  "type": "LLM_MODEL",
  "version": {
    "name": "v1",
    "description": "test description",
    "config": {
      "llm_model_config": {
        "base_model_type": "",
        "model_path": ""
      }
    }
  }
}
```

响应示例:

```
{
  "id": "ac8111bf-3601-4905-8ddd-b41d3e636a4e"
}
```

4. 调用创建推理服务接口，创建一个推理服务，记录接口返回推理服务的ID。

请求示例：

POST https://{hostname}/v1/workspaces/{workspace_id}/services/instances

workspace_id：为步骤1中记录的工作空间ID。

Body：

```
{
  "source": {
    "id": ""
  },
  "name": "test_serviceInstanceName",
  "description": "description",
  "endpoint_id": ""
}
```

- id：为步骤3中接口返回记录的模型ID。
- endpoint_id：为步骤2中接口返回记录的推理端点ID。

响应示例：

```
{
  "id": "b935d0ef-f4eb-4b95-aff1-9d33ae9f57b6"
}
```

5. 调用推理请求接口，发起推理请求。

请求示例：

POST https://{hostname}/v1/workspaces/{workspace_id}/services/instances/{instance_id}/invocations

- workspace_id：为步骤1中记录的工作空间ID。
- instance_id：为步骤4中记录的推理服务的ID。

Body：

```
{
  "messages": [
    {
      "role": "user",
      "content": "hello"
    }
  ]
}
```

响应示例：推理请求接口为流式返回。

```
{
  "id": "chatcmpl-62dda7304f53451c9477e0",
  "object": "chat.completion.chunk",
  "created": 1730120529,
  "model": "ada1d67d-f2a1-4e77-838f-0d8688d756f4",
  "choices": [
    {
      "index": 0,
      "delta": {
        "role": "assistant",
        "content": "\n\nHello! LLM stands for Large Language Model. It refers to artificial intelligence models, like myself,"
      },
      "finish_reason": null
    }
  ],
  "system_fingerprint": null,
  "usage": null
}
```

6 权限和授权项

6.1 权限及授权项说明

如果您需要对您所拥有的DataArtsFabric进行精细的权限管理，您可以使用统一身份认证服务（Identity and Access Management，简称IAM），如果华为云账号已经能满足您的要求，不需要创建独立的IAM用户，您可以跳过本章节，不影响您使用DataArtsFabric服务的其它功能。

默认情况下，新建的IAM用户没有任何权限，您需要将其加入用户组，并给用户组授予策略或角色，才能使用户组中的用户获得相应的权限，这一过程称为授权。授权后，用户就可以基于已有权限对云服务进行操作。

权限根据授权的精细程度，分为**角色**和**策略**。角色以服务为粒度，是IAM最初提供了一种根据用户的工作职能定义权限的粗粒度授权机制。策略以API接口为粒度进行权限拆分，授权更加精细，可以精确到某个操作、资源和条件，能够满足企业对权限最小化的安全管控要求。

说明

如果您要允许或是禁止某个接口的操作权限，请使用策略。

账号具备所有接口的调用权限，如果使用账号下的IAM用户发起API请求时，该IAM用户必须具备调用该接口所需的权限，否则，API请求将调用失败。每个接口所需要的权限，与各个接口所对应的授权项相对应，只有发起请求的用户被授予授权项所对应的策略，该用户才能成功调用该接口。例如，用户要调用接口来查询云服务器列表，那么这个IAM用户被授予的策略中必须包含允许“ecs:servers:list”的授权项，该接口才能调用成功。

支持的授权项

策略包含系统策略和自定义策略，如果系统策略不满足授权要求，管理员可以创建自定义策略，并通过给用户组授予自定义策略来进行精细的访问控制。策略支持的操作与API相对应，授权项列表说明如下：

- 权限：允许或拒绝某项操作。
- 对应API接口：自定义策略实际调用的API接口。
- 授权项：自定义策略中支持的Action，在自定义策略中的Action中写入授权项，可以实现授权项对应的权限功能。

- 依赖的授权项：部分Action存在对其他Action的依赖，需要将依赖的Action同时写入授权项，才能实现对应的权限功能。
- IAM项目(Project)/企业项目(Enterprise Project)：自定义策略的授权范围，包括IAM项目与企业项目。授权范围如果同时支持IAM项目和企业项目，表示此授权项对应的自定义策略，可以在IAM和企业管理两个服务中给用户组授权并生效。如果仅支持IAM项目，不支持企业项目，表示仅能在IAM中给用户组授权并生效，如果在企业管理中授权，则该自定义策略不生效。关于IAM项目与企业项目的区别，详情请参见[IAM与企业管理的区别](#)。

6.2 策略授权参考

本章节介绍DataArtsFabric策略授权场景下支持的策略授权项。

支持的授权项

策略包含系统策略和自定义策略，如果系统策略不满足授权要求，管理员可以创建自定义策略，并通过给用户组授予自定义策略来进行精细的访问控制。策略支持的操作与API相对应，授权项列表说明如下：

- 权限：允许或拒绝某项操作。
- 对应API接口：自定义策略实际调用的API接口。
- 授权项：自定义策略中支持的Action，在自定义策略中的Action中写入授权项，可以实现授权项对应的权限功能。
- 依赖的授权项：部分Action存在对其他Action的依赖，需要将依赖的Action同时写入授权项，才能实现对应的权限功能。
- IAM项目（Project）/企业项目（Enterprise Project）：自定义策略的授权范围，包括IAM项目与企业项目。授权范围如果同时支持IAM项目和企业项目，表示此授权项对应的自定义策略，可以在IAM和企业管理两个服务中给用户组授权并生效。如果仅支持IAM项目，不支持企业项目，表示仅能在IAM中给用户组授权并生效，如果在企业管理中授权，则该自定义策略不生效。管理员可以在授权项列表中查看授权项是否支持IAM项目或企业项目，“√”表示支持，“×”表示暂不支持。关于IAM项目与企业项目的区别，详情请参见[IAM与企业管理的区别](#)。

DataArtsFabric的支持自定义策略授权项如下所示：

生命周期管理

表 6-1 生命周期管理

权限	对应API接口	授权项（Action）	依赖的授权项	IAM项目（Project）	企业项目（Enterprise Project）
查询Workspace列表	GET /v1/workspaces	DataArtsFabric:workspace:list	-	√	√

权限	对应API接口	授权项 (Action)	依赖的授权项	IAM 项目 (Proj ect)	企业项 目 (Enter prise Project)
创建 Worksp ace	POST /v1/ workspaces	DataArtsFa bric:worksp ace:create	lakeformation:in stance:describe	√	√
修改 Worksp ace	PUT /v1/workspaces/ {workspace_id}	DataArtsFa bric:worksp ace:alter	lakeformation:in stance:describe	√	×
修改 Worksp ace监控 配置	PUT /v1/workspaces/ {workspace_id}/ metrics	DataArtsFa bric:worksp ace:alterMe trics	-	√	×
删除 Worksp ace	DELETE /v1/workspaces/ {workspace_id}	DataArtsFa bric:worksp ace:drop	-	√	×
查询计 算资源	GET /v1/workspaces/ {workspace_id}/ computes	DataArtsFa bric:worksp ace:listCom pute	-	√	×
创建计 算资源	POST /v1/workspaces/ {workspace_id}/ computes	DataArtsFa bric:worksp ace:createC ompute	-	√	×
修改计 算资源	PUT /v1/ workspaces/ {workspace_id}/ computes/ {compute_id}	DataArtsFa bric:worksp ace:alterCo mpute	-	√	×
删除计 算资源	DELETE /v1/ workspaces/ {workspace_id}/ computes/ {compute_id}	DataArtsFa bric:worksp ace:dropCo mpute	-	√	×
查询 Worksp ace的 Endpoi nt列表	GET /v1/workspaces/ {workspace_id}/ endpoints	DataArtsFa bric:endpoi nt:list	-	√	×

权限	对应API接口	授权项 (Action)	依赖的授权项	IAM 项目 (Proj ect)	企业项 目 (Enter prise Project)
创建 Worksp ace的 Endpoi nt	POST /v1/workspaces/ {workspace_id}/ endpoints	DataArtsFa bric:endpoi nt:create	-	√	×
查询 Worksp ace的 Endpoi nt详情	GET /v1/ workspaces/ {workspace_id}/ endpoints/ {endpoint_id}	DataArtsFa bric:endpoi nt:show	-	√	×
修改 Worksp ace的 Endpoi nt	PUT /v1/ workspaces/ {workspace_id}/ endpoints/ {endpoint_id}	DataArtsFa bric:endpoi nt:alter	-	√	×
删除 Worksp ace的 Endpoi nt	DELETE /v1/ workspaces/ {workspace_id}/ endpoints/ {endpoint_id}	DataArtsFa bric:endpoi nt:drop	DataArtsFabric:jo b:listJobInstance DataArtsFabric:s ervice:listInstanc e	√	×
订阅公 共端点	POST /v1/ workspaces/ {workspace_id}/ endpoints/ {endpoint_id}/ subscribe	DataArtsFa bric:endpoi nt:subscribe	-	√	×
查询作 业列表	GET /v1/workspaces/ {workspace_id}/ jobs	DataArtsFa bric:job:list	-	√	×
创建作 业	POST /v1/workspaces/ {workspace_id}/ jobs	DataArtsFa bric:job:crea te	-	√	×
查询作 业	GET /v1/ workspaces/ {workspace_id}/ jobs/{job_id}/ versions	DataArtsFa bric:job:sho w	-	√	×

权限	对应API接口	授权项 (Action)	依赖的授权项	IAM 项目 (Proj ect)	企业项 目 (Enter prise Project)
修改作业	PUT /v1/ workspaces/ {workspace_id}/ jobs/{job_id}	DataArtsFa bric:job:alte r	-	√	×
删除作业	DELETE /v1/ workspaces/ {workspace_id}/ jobs/{job_id}	DataArtsFa bric:job:dro p	DataArtsFabric:jo b:listJobInstance	√	×
创建模型	POST /v1/workspaces/ {workspace_id}/ models	DataArtsFa bric:model:c reate	obs:bucket:Head Bucket obs:bucket:ListBu cketVersions obs:bucket:ListAll MyBuckets obs:bucket:ListBu cket	√	×
查询模型列表	GET /v1/ workspaces/ {workspace_id}/ models	DataArtsFa bric:model:l ist	-	√	×
查询模型	GET /v1/ workspaces/ {workspace_id}/ models/ {model_id}/ versions	DataArtsFa bric:model:s how	-	√	×
删除模型	DELETE /v1/ workspaces/ {workspace_id}/ models/ {model_id}/ versions	DataArtsFa bric:model: drop	DataArtsFabric:s ervice:listInstanc e	√	×
修改模型	PUT /v1/ workspaces/ {workspace_id}/ models/ {model_id}/ versions	DataArtsFa bric:model: alter	-	√	×

权限	对应API接口	授权项 (Action)	依赖的授权项	IAM 项目 (Proj ect)	企业项 目 (Enter prise Project)
创建标 签	POST /v1/{project_id}/ fabric-workspace/ {workspace_id}/ tags/create	DataArtsFa bric:worksp ace:tagReso urce	-	√	×
删除标 签	POST /v1/{project_id}/ fabric-workspace/ {workspace_id}/ tags/delete	DataArtsFa bric:worksp ace:unTagR esource	-	√	×
获取标 签列表	GET /v1/{project_id}/ fabric-workspace/ tags	DataArtsFa bric:worksp ace:listTags	-	√	×
查询指 定资源 标签	GET /v1/ {project_id}/fabric- workspace/ {workspace_id}/ tags	DataArtsFa bric:worksp ace:listTags ForResource	-	√	×
标签查 询资源 列表	POST /v1/ {project_id}/fabric- workspace/ resource- instances/filter	DataArtsFa bric:worksp ace:listReso urcesByTag	-	√	√
创建消 息通知 策略	POST /v1/workspaces/ {workspace_id}/ messages	DataArtsFa bric:worksp ace:createM essagePolic y	iam:agencies:list Agencies iam:roles:listRole s iam:permissions:l istRolesForAgenc y smn:topic:list	√	×
查询消 息通知 策略列 表	GET /v1/workspaces/ {workspace_id}/ messages	DataArtsFa bric:worksp ace:listMess agePolicy	-	√	×

权限	对应API接口	授权项 (Action)	依赖的授权项	IAM 项目 (Proj ect)	企业项 目 (Enter prise Project)
删除消息通知策略	DELETE /v1/ workspaces/ {workspace_id}/ messages/ {message_policy_id }	DataArtsFa bric:worksp ace:deleteM essagePolic y	-	√	×
查询运行作业列表	GET /v1/workspaces/ {workspace_id}/ jobs/runs	DataArtsFa bric:job:list JobInstance	DataArtsFabric:w orkspace:list	√	×
运行作业	POST /v1/workspaces/ {workspace_id}/ jobs/runs	DataArtsFa bric:job:run	DataArtsFabric:w orkspace:list DataArtsFabric:e ndpoint:show	√	×
查询运行作业	GET /v1/ workspaces/ {workspace_id}/ jobs/runs/{run_id}	DataArtsFa bric:job:sho wJobInstan ce	DataArtsFabric:w orkspace:list DataArtsFabric:e ndpoint:show	√	×
删除运行作业	DELETE /v1/ workspaces/ {workspace_id}/ jobs/runs/{run_id}	DataArtsFa bric:job:dro pJobInstanc e	DataArtsFabric:w orkspace:list DataArtsFabric:e ndpoint:show	√	×
取消运行作业	POST /v1/ workspaces/ {workspace_id}/ jobs/runs/{run_id}/ cancel	DataArtsFa bric:job:can celJobInsta nce	DataArtsFabric:w orkspace:list DataArtsFabric:e ndpoint:show	√	×
调用推理服务实例	POST /v1/ workspaces/ {workspace_id}/ services/instances/ {instance_id}/ invocations	DataArtsFa bric:model:i nference	DataArtsFabric:w orkspace:list DataArtsFabric:e ndpoint:show	√	×
查询路由列表	GET /v1/workspaces/ {workspace_id}/ routes	DataArtsFa bric:worksp ace:listRout e	DataArtsFabric:w orkspace:list	√	×

权限	对应API接口	授权项 (Action)	依赖的授权项	IAM 项目 (Proj ect)	企业项 目 (Enter prise Project)
查询 session 信息	POST /v1/sessions	DataArtsFa bric:worksp ace:showSe ssion	-	√	×

7 附录

7.1 状态码

状态码如下表所示。

表 7-1 状态码

状态码	编码	错误码说明
100	Continue	继续请求。 这个临时响应用来通知客户端，它的部分请求已经被服务器接收，且仍未被拒绝。
101	Switching Protocols	切换协议。只能切换到更高级的协议。 例如，切换到HTTP的新版本协议。
200	OK	接口调用成功。
201	Created	创建类的请求完全成功。
202	Accepted	已经接受请求，但未处理完成。
203	Non-Authoritative Information	非授权信息，请求成功。
204	NoContent	请求完全成功，同时HTTP响应不包含响应体。 在响应OPTIONS方法的HTTP请求时返回此状态码。
205	Reset Content	重置内容，服务器处理成功。
206	Partial Content	服务器成功处理了部分GET请求。
300	Multiple Choices	多种选择。请求的资源可包括多个位置，相应可返回一个资源特征与地址的列表用于用户终端（例如：浏览器）选择。

状态码	编码	错误码说明
301	Moved Permanently	永久移动，请求的资源已被永久的移动到新的URI，返回信息会包括新的URI。
302	Found	资源被临时移动。
303	See Other	查看其它地址。使用GET和POST请求查看。
304	Not Modified	所请求的资源未修改，服务器返回此状态码时，不会返回任何资源。
305	Use Proxy	所请求的资源必须通过代理访问。
306	Unused	已经被废弃的HTTP状态码。
400	BadRequest	非法请求。 建议直接修改该请求，不要重试该请求。
401	Unauthorized	在客户端提供认证信息后，返回该状态码，表明服务端指出客户端所提供的认证信息不正确或非法。
402	Payment Required	保留请求。
403	Forbidden	请求被拒绝访问。 返回该状态码，表明请求能够到达服务端，且服务端能够理解用户请求，但是拒绝做更多的事情，因为该请求被设置为拒绝访问，建议直接修改该请求，不要重试该请求。
404	NotFound	所请求的资源不存在。 建议直接修改该请求，不要重试该请求。
405	MethodNotAllowed	请求中带有该资源不支持的方法。 建议直接修改该请求，不要重试该请求。
406	Not Acceptable	服务器无法根据客户端请求的内容特性完成请求。
407	Proxy Authentication Required	请求要求代理的身份认证，与401类似，但请求者应当使用代理进行授权。
408	Request Time-out	服务器等候请求时发生超时。 客户端可以随时再次提交该请求而无需进行任何更改。
409	Conflict	服务器在完成请求时发生冲突。 返回该状态码，表明客户端尝试创建的资源已经存在，或者由于冲突请求的更新操作不能被完成。
410	Gone	客户端请求的资源已经不存在。 返回该状态码，表明请求的资源已被永久删除。
411	Length Required	服务器无法处理客户端发送的不带Content-Length的请求信息。

状态码	编码	错误码说明
412	Precondition Failed	未满足前提条件，服务器未满足请求者在请求中设置的其中一个前提条件。
413	Request Entity Too Large	由于请求的实体过大，服务器无法处理，因此拒绝请求。为防止客户端的连续请求，服务器可能会关闭连接。如果只是服务器暂时无法处理，则会包含一个Retry-After的响应信息。
414	Request-URI Too Large	请求的URI过长（URI通常为网址），服务器无法处理。
415	Unsupported Media Type	服务器无法处理请求附带的媒体格式。
416	Requested range not satisfiable	客户端请求的范围无效。
417	Expectation Failed	服务器无法满足Expect的请求头信息。
422	Unprocessable Entity	请求格式正确，但是由于含有语义错误，无法响应。
429	TooManyRequests	表明请求超出了客户端访问频率的限制或者服务端接收到多于它能处理的请求。建议客户端读取相应的Retry-After首部，然后等待该首部指出的时间后再重试。
500	InternalServerError	表明服务端能被请求访问到，但是不能理解用户的请求。
501	Not Implemented	服务器不支持请求的功能，无法完成请求。
502	Bad Gateway	充当网关或代理的服务器，从远端服务器接收到了一个无效的请求。
503	Service Unavailable	被请求的服务无效。 建议直接修改该请求，不要重试该请求。
504	Server Timeout	请求在给定的时间内无法完成。客户端仅在为请求指定超时（Timeout）参数时会得到该响应。
505	HTTP Version not supported	服务器不支持请求的HTTP协议的版本，无法完成处理。

7.2 错误码

调用接口出错后，将不会返回结果数据。调用方可根据每个接口对应的错误码来定位错误原因。当调用出错时，HTTP请求返回一个4xx或5xx的HTTP状态码。返回的消息体中是具体的错误代码及错误信息。在调用方找不到错误原因时，可以联系华为云客服，并提供错误码，以便我们尽快帮您解决问题。

错误响应 Body 体格式说明

当接口调用出错时，会返回错误码及错误信息说明，错误响应的Body体格式如下所示。

```
{
  "error_msg": "Inference features are not enabled.",
  "error_code": "dataartsfabric.00010195"
}
```

其中，error_code表示错误码，error_msg表示错误描述信息。

错误码说明

当您调用API时，如果遇到“APIGW”开头的错误码，请参见[API网关错误码](#)进行处理。

表 7-2 DataArtsFabric 错误码

状态码	错误码	错误信息	处理措施
500	00000001	运行中，请稍后。	请休息下，重新获取。
400	00010003	工作空间不存在。	请先创建工作空间。
400	00010004	输入的桶不存在。	请检查输入的桶。
400	00010005	输入的协议版本不存在。	请检查输入的协议版本。
400	00010006	输入的协议版本不能为空。	请检查输入的协议版本。
400	00010007	资源清理任务id:xx错误。	请检查任务ID。
400	00010009	查询的资源不存在。	请输入正确的资源ID和项目ID。
400	00010011	该规格编码是无效的。	请输入正确的规格编码。
400	00010022	该特性【xx】为受限特性，面向白名单客户开放。	请联系运维人员申请白名单。
400	00010023	参数中的时间格式非法。	请使用合法的时间格式。
400	00010024	服务委托异常。	请在服务授权页面再次同意服务授权。
400	00010026	标签列表中的key重复	请检查标签列表中的键。
400	00010028	标签列表中成员的值列表中存在重复的值。	请检查标签列表中成员的值列表。
400	00010029	标签列表中的成员不能为NULL。	请检查标签列表中的成员。

状态码	错误码	错误信息	处理措施
400	00010030	条件列表中的成员不能为NULL。	请检查条件列表中的成员。
400	00010031	标签列表中成员的值列表不能包含NULL。	请检查标签列表中成员的值列表。
400	00010032	标签列表中成员的值列表不能为NULL。	请检查标签列表中成员的值列表。
400	00010033	条件列表成员的值不能为NULL。	请检查条件列表成员的值。
400	00010034	当前不支持批量删除系统标签。	请选择其他标签操作。
400	00010035	您没有权限操作系统标签。	请检查您的权限。
400	00010036	系统标签不能为空。	请检查系统标签参数。
400	00010037	系统标签和标签不能同时使用。	请选择系统标签或者标签中的一种。
400	00010038	请求未被授权。	请检查相关权限：xx，先授权再请求。
400	00010043	资源不足，无法创建工作空间。	请联系运维人员检查资源。
400	00010046	该接口达到限流阈值。	请稍后重试。
400	00010049	参数不合法，参数值是xx。	请检查参数是否合法。
400	00010051	没有发送消息权限。	请确认配置是否正确。
400	00010058	在过载控制中收集不到账号ID。	请检查账号信息。
400	00010059	接入租户未授权。	请检查接入租户是否已经授权。
400	00010060	接入租户已授权。	请不要重复授权接入租户。
400	00010062	当前任务状态下不能进行此操作。	请检查任务状态。
400	00010063	xx已存在。	请检查输入的任务名xx。
400	00010064	xx数量已达上限。	请删除空闲的任务后重新创建或联系运维人员提高数量上限。
400	00010065	风险检查失败。	请检查您的余额。
400	00010066	任务xx不存在。	请检查Job Id。

状态码	错误码	错误信息	处理措施
400	00010067	初始化任务失败，可能已达任务上限。	请稍后重试。
400	00010068	应用当前状态为xx不能被删除，状态为DELETING或者DELETED的应用不能再次被删除。	当前应用已经是删除中或者已删除状态，无需再次删除。
400	00010069	应用不存在。	请使用正确的应用ID。
400	00010071	没有找到该商品。规格编码:xx。	请购买已有的商品。
400	00010072	没有找到商品ID。	请重试。
400	00010073	没有权限购买xx规格。	请购买其他商品。
400	00010075	应用当前状态为xx不能被更新，状态为STOPPED的应用才能被更新。	请等待应用状态为STOPPED之后再更新。
400	00010076	应用不存在。	请使用正确的应用ID。
400	00010077	应用版本不存在。	请使用正确的版本ID。
400	00010081	创建应用失败。错误原因：xx。	请根据失败原因重试。
400	00010082	日志文件不存在。	请查看您配置的日志存储路径下是否有文件存在。
400	00010083	作业没有运行过。	请查看确保您的作业运行过，才能查看日志。
400	00010084	任务正在运行中。	请检查任务状态，稍后重试。
400	00010085	任务未在运行中。	请检查任务状态，稍后重试。
400	00010086	xxQPS规格不存在。	请检查QPS规格。
400	00010089	桶对象不存在。	请检查OBS地址。
400	00010090	YAML文件格式错误。	请检查YAML文件的格式。
400	00010091	许可证不存在，区域：xx，esn：xx。	请检查许可证是否存在。
400	00010092	无匹配许可证模式：xx。	请检查许可证模式。
400	00010093	许可证bbom为空。	请检查许可证bbom。
400	00010094	无匹配bbom编号：xx。	请检查bbom编号。
400	00010095	无匹配bbom类型：xx。	请检查bbom类型。
400	00010096	许可证解析失败。	请检查许可证文件内容。

状态码	错误码	错误信息	处理措施
400	00010097	许可证超量。	请检查许可证。
400	00010098	许可证即将超量。	请检查许可证。
400	00010099	无匹配的许可证内容， lsn: xx	请检查许可证文件内容。
400	00010100	无匹配bbom，区域： xx。	请检查许可证bbom。
400	00010101	许可证内容esn不匹配： xx，许可证esn: xx。	请检查许可证内容esn编号。
400	00010102	许可证内容已过期。	请检查许可证内容是否过期。
400	00010103	任务资源被占用。	请等待重试。
400	00010104	客户端接入模式不是自动 创建模式。	请检查客户端接入模式。
400	00010105	创建应用版本失败，错误 原因: xx。	请根据失败原因重试。
400	00010107	任务在未被授权的实例xx 中。	请用户在xx实例被授权后再 试。
400	00010108	任务在不存在的实例xx 中。	请用户恢复被删除的xx实例， 或在已有实例重新创建任务后 再试。
400	00010110	SMN消息校验失败。	请检查消息内容是否被篡改。
400	00010111	任务参数xx不存在。	请创建或修改任务使其包含参 数xx。
400	00010116	输入的qps规格参数为 空。	请检查输入的参数。
400	00010117	只需要一个qps规格参 数。	请检查输入的参数。
400	00010118	特性目前暂不支持。	特性暂不支持，请等待版本更 新。
400	00010120	数据源类型xx不支持。	请检查数据源类型。
400	00010121	权限策略类型xx不支持。	请检查权限策略类型。
400	00010122	输入的VPC网段与 Lakeformation网段冲 突。	请检查输入的VPC网段。
400	00010123	元数据迁移网络类型错 误。	请检查迁移网络类型。
400	00010124	SMN服务传入了无效的 确认订阅或证书URL。	请传入HTTPS类型的URL。

状态码	错误码	错误信息	处理措施
400	00010125	DLF接入点错误。	请检查DLF接入点。
400	00010127	许可证的API访问请求控制项的容量不足。	请检查许可证的API访问请求控制项的容量。
400	00010128	迁移对象为空。	请检查任务中的迁移对象参数。
400	00010130	xx不存在。	请使用正确的条件。
400	00010132	xx已存在。	请检查输入。
400	00010133	注册CAP失败，xx。	请确认CAP的信息是否正确。
400	00010134	查询的用户不存在。	请输入正确的用户信息。
400	00010135	集群ID或Session不正确。	请输入正确的信息。
400	00010136	子网数量超出限制。	请检查输入中的子网数量。
400	00010137	删除CAP失败，xx。	请检查CAP共享状态以及设置CAP当前版本为空。
400	00010138	创建Endpoint部署实例失败。错误原因：xx。	请参考日志定位失败原因或联系运维人员。
400	00010142	创建服务委托失败：xx。	请检查当前租户委托数量上限，或申请配额上限后重试。
400	00010143	xx资源不足，无法创建。	请检查您剩余的资源，确保您使用的资源小于剩余资源。
400	00010144	当前租户还拥有Workspace，无法取消授权。	请将本租户拥有的Workspace全部删除后，再取消授权。
400	00010145	类型错误，xx。	请输入正确的类型。
400	00010146	当前版本无法删除。	请等待版本不是当前版本的时候再删除。
400	00010147	没有资源访问权限。	请检查您访问资源的权限。
400	00010148	节点池不存在。	请检查输入。
400	00010149	不能删除共享的App。	请确认APP是非共享状态再删除。
400	00010150	请求的Endpoint不允许执行该操作。	请确认操作的不是默认Endpoint，并且按需Endpoint只有name或description字段更新。
400	00010151	输入参数不合法，xx。	请输入正确的参数值：xx。

状态码	错误码	错误信息	处理措施
400	00010152	不能删除存在Job运行实例或Service实例的Endpoint。	请删除当前Endpoint下的Job运行实例和服务实例后，再重试。
400	00010153	当前Endpoint状态下不能进行此操作。	请检查Endpoint状态。
400	00010154	不支持的注册资源类型：xx。	支持以下资源类型：xx。
400	00010155	当前Workspace下还拥有xx资源，无法删除。	请将当前Workspace下xx资源全部删除后，再删除Workspace。
400	00010157	要授权的obs路径已授权：xx。	请检查输入的obs路径参数。
400	00010158	OBS服务发生异常，OBS请求responseCode：xx，请求ID：xx。	请检查OBS服务状态和OBS路径：xx。
400	00010159	要授权的obs路径不存在或不是并行文件系统：xx。	请检查输入的obs路径参数。
400	00010162	工作空间名称已经存在。	请输入不同的工作空间名称。
400	00010163	不支持的策略类型：xx。	请参考官网的API文档输入正确的策略类型。
400	00010164	不能删除必需的权限策略：xx。	请检查输入。
400	00010165	不支持的消息类型：xx。	请参考官网的API文档输入正确的消息类型。
400	00010166	不支持的通知类型：xx。	请参考官网的API文档输入正确的通知类型。
400	00010167	没有授权SMN委托权限策略，无法使用消息通知功能。	请在服务授权界面授权SMN权限：xx。
400	00010168	当前工作空间消息通知策略数量已达上限：xx。	请先删除多余的消息通知策略再创建新的策略。
400	00010169	参数中存在相同的消息通知策略：xx。	请检查输入。
400	00010170	没有找到消息通知策略，策略ID：xx。	请输入已存在策略的ID。
400	00010171	OBS路径不存在：xx。	请检查OBS地址。
400	00010172	不支持的包周期回调操作类型xx。	请检查输入的操作类型。

状态码	错误码	错误信息	处理措施
400	00010173	不支持的企业项目IDxx。	请检查输入的企业项目ID。
400	00010174	当前app正在使用无法删除。	请等待App未被使用的时候再删除。
400	00010175	不能删除共享App的版本。	请确认App是非共享状态再删除版本。
400	00010176	当前Domain下工作空间创建数量已达到上限。	请删除空闲的工作空间后重新创建或联系运维人员提高数量上限。
400	00010177	CAP名称xx在当前工作空间xx已存在。	请使用其它CAP名称。
400	00010178	CAP版本名称xx在当前CAP中已存在。	请使用其它CAP版本名称。
400	00010179	操作App实例异常：xx。	请检查输入。
400	00010180	实例冻结不可释放。	请在实例解冻后删除。
400	00010181	冻结实例变更受限，仅允许修改基本信息。	请在实例解冻后操作。
400	00010183	请求的资源超限。	请检查最大资源限制配置。
400	00010185	该特性暂未开放。	请联系运维确认特性能力集。
400	00010188	Metastore首次添加后不可修改。	请检查Metastore是否发生改变。
400	00010189	有xx使用当前版本，无法删除。	请删除所有使用该版本的xx后再重试。
400	00010190	Job相关特性未开通。	请联系运维人员开通Job特性。
400	00010191	Ray相关特性未开通。	请联系运维人员开通Ray特性。
400	00010192	暂不支持修改端点的资源规格。	端点的资源规格暂不支持修改。
400	00010193	暂不支持的请求参数值{ }。	请检查输入。
400	00010194	当前Workspace下创建的Endpoint数量已达到上限。	请删除空闲的Endpoint后重新创建或联系运维人员提高数量上限。
400	00010195	推理相关特性未开通。	请联系运维人员开通推理特性。
400	00010196	Ray集群状态异常。	请删除后重新创建或联系运维人员。
400	00010197	计算资源状态非运行中。	请恢复计算资源状态或删除计算资源后新建一个。

状态码	错误码	错误信息	处理措施
400	00090002	在过载控制中收集不到账号ID。	请检查账号信息。
400	00090003	申请资源组时资源不足。	请扩容Workspace中的资源。
400	00090004	找不到Workspace。	请选择正确的Workspace。
400	00090005	找不到Endpoint。	请选择正确的Endpoint。
400	00090006	找不到Workspace中对应的资源。	请确认Workspace中购买了对应的资源。
400	00090008	创建Endpoint失败。错误原因：xx。	请根据失败原因重试。
400	00090009	xx不存在	请选择正确的信息。
400	00090012	没权限使用：xx。	请检查是否有权限。
400	00090013	作业运行失败：xx。	请检查运行作业的配置信息。
400	00090014	无法删除活动中的任务：xx。	请等待任务运行完成或者停止正在运行的作业。
400	00090015	无法删除公共实例：xx。	请检查实例是否是公共的。
400	00090016	停止服务失败：xx。	请确认该实例是否可以取消。
400	00090017	停止任务失败：xx。	请检查该任务的状态。
400	00090018	更新任务失败：xx。	请确认该任务是否可以升级。
400	00090019	创建服务失败：xx。	请检查输入参数。
400	00090020	删除服务失败：xx。	请检查输入参数。
400	00090021	更新服务失败：xx。	请检查更新时的输入参数。
400	00090022	输入参数不合法，xx。	请输入正确的参数值：xx。
400	00090023	推理失败：xx。	请输入正确的推理请求：xx。
400	00090024	请求的Endpoint不允许执行该操作。	请确认操作的是按需Endpoint。
400	00090025	模型基础类型xx与Endpoint资源不匹配。	请检查app的基础模型类型是否和Endpoint资源相匹配。
400	00010038	请求未被授权。	请检查相关权限：xx，先授权再请求。
400	00090047	OBS文件下载异常xx。	请检查OBS下载配置信息xx。
400	00090050	OBS非法路径异常xx。	请输入正确的OBS路径xx。
400	00090058	请求的资源超限。	请检查最大资源限制配置。
400	00090059	请求的Token为空	请携带Token访问。

状态码	错误码	错误信息	处理措施
400	00090060	APP实例副本已经存在于指定的引擎实例。	请检查输入的APP实例副本ID是否重复或错误。
400	00090061	此方法未实现。	请勿使用此方法。
400	00090062	端点维护中。	请稍后重试或在其他端点上创建APP实例。
400	00090063	端点冻结中。	请尝试恢复端点状态或者联系运维人员。
400	00090064	端点未开通。	使用前请先开通公共端点。
400	00090067	推理相关特性未开通。	请联系运维人员开通推理特性。

7.3 获取项目 ID

调用 API 获取项目 ID

项目ID还可以通过调用[查询指定条件下的项目列表](#)API获取。

获取项目ID的接口为“GET https://{Endpoint}/v3/projects”，其中{Endpoint}为IAM的终端节点，可以从[地区和终端节点](#)获取。接口的认证鉴权请参见[认证鉴权](#)。

响应示例如下，其中projects下的“id”即为项目ID。

```
{
  "projects": [
    {
      "domain_id": "65382450e8f64ac0870cd180d14e****",
      "is_domain": false,
      "parent_id": "65382450e8f64ac0870cd180d14e****",
      "name": "cn-north-4",
      "description": "",
      "links": {
        "next": null,
        "previous": null,
        "self": "https://www.example.com/v3/projects/a4a5d4098fb4474fa22cd05f897d****"
      },
      "id": "a4a5d4098fb4474fa22cd05f897d****",
      "enabled": true
    }
  ],
  "links": {
    "next": null,
    "previous": null,
    "self": "https://www.example.com/v3/projects"
  }
}
```

从控制台获取项目 ID

在调用接口的时候，部分URL中需要填入项目编号，所以需要获取到项目编号。项目编号获取步骤如下：

1. 登录管理控制台。
2. 鼠标移动到右上角的用户名处，在下拉列表中单击“我的凭证”。
3. 在“API凭证”页面的项目列表中查看项目ID。
多项目时，展开“所属区域”，从“项目ID”列获取子项目ID。

7.4 获取账号 ID

在调用接口的时候，部分URL中需要填入账号ID，所以需要先在管理控制台上获取到账号ID。账号ID获取步骤如下：

1. 登录管理控制台。
2. 鼠标移动到右上角的用户名处，在下拉列表中单击“我的凭证”。
3. 在“API凭证”页面中查看账号ID。

7.5 获取工作空间 ID

1. 登录DataArtsFabric工作空间管理台，在目标工作空间卡片单击“查看详情”。

图 7-1 查看详情



2. 在“基础信息”页面获取当前工作空间的ID。

图 7-2 获取工作空间 ID

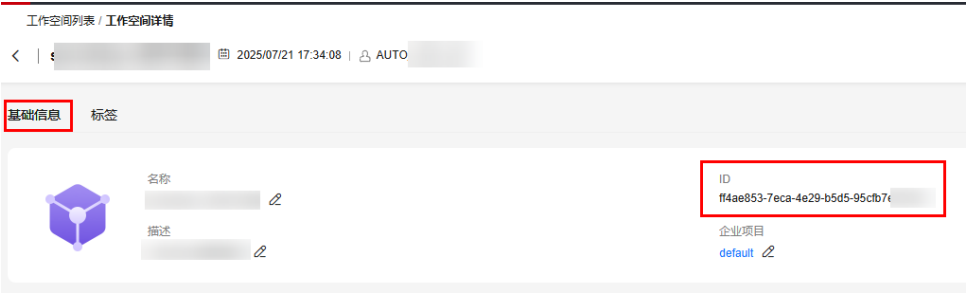
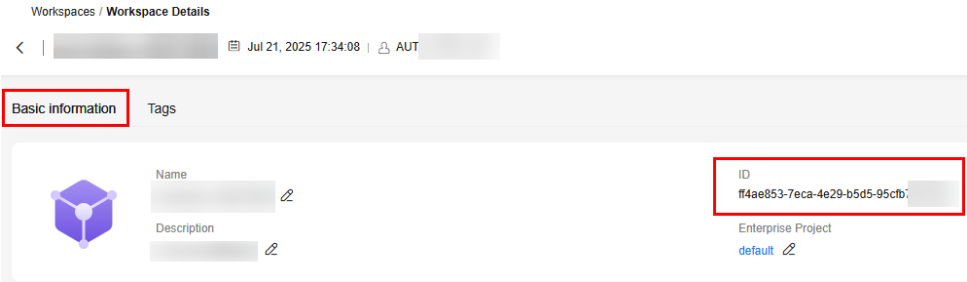


图 7-3 英文图



7.6 获取端点 ID

1. 登录DataArtsFabric工作空间管理台，在目标工作空间卡片单击“进入工作空间”。

图 7-4 进入工作空间



2. 在左侧导航栏，单击“推理端点”，按需选择“我的端点”或“公共端点”页签。
3. 单击目标端点卡片，在“概览”页面查看端点ID。

图 7-5 获取端点 ID



7.7 获取模型 ID

- 步骤1** 登录DataArtsFabric工作空间管理台。
- 步骤2** 选择已创建的工作空间，单击“进入工作空间”，选择“资源与资产 > 模型”。
- 步骤3** 按需选择“我的模型”或“公共模型”页签，单击模型名称。
- 步骤4** 在“模型详情”页面，获取模型ID。

图 7-6 获取模型 ID



----结束

7.8 获取推理服务实例 ID

- 步骤1** 登录DataArtsFabric工作空间管理台。
- 步骤2** 选择已创建的工作空间，单击“进入工作空间”，选择“开发与生产 > 推理服务”。
- 步骤3** 按需选择“我的推理服务”或“公共推理服务”页签，单击推理服务名称。
- 步骤4** 在“概览”页面，获取推理实例ID。

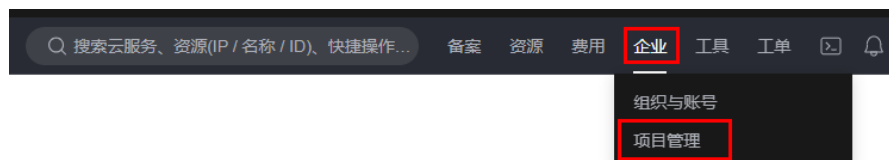
----结束

7.9 获取企业项目 ID

步骤1 登录DataArtsFabric工作空间管理台。

步骤2 在顶部导航栏，选择“企业 > 项目管理”。

图 7-7 项目管理



步骤3 在“企业项目管理”页面的“名称/ID”列，单击复制按钮，即可复制企业项目ID。

图 7-8 获取企业项目 ID



----结束