

对象存储

元数据与数据发现



腾讯云

【 版权声明 】

©2013–2025 腾讯云版权所有

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

【 商标声明 】



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

【 服务声明 】

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

【 联系我们 】

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

文档目录

元数据与数据发现

元数据与数据发现概述

管理文件元数据

数据集管理

创建数据集

查询数据集

更新数据集

删除数据集

列出数据集

绑定存储桶与数据集

解绑存储桶与数据集

查询数据集与存储桶的绑定关系

查询绑定关系列表

检索模板与算子

数据索引

数据索引概述

索引操作

创建元数据索引

更新元数据索引

查询元数据索引

删除元数据索引

标量检索

标量检索概述

简单查询

查询结果异步导出

创建异步导出任务

查询异步导出任务

列出异步导出任务

字段和操作符的支持列表

向量检索

向量检索概述

图像检索

人脸检索

元数据与数据发现

元数据与数据发现概述

最近更新时间：2025-02-17 18:27:12

产品简介

智能检索 MetaInsight 为您提供多场景、跨模态的检索服务，您可以通过自然语言或结构化的检索条件，分析存储在对象存储 COS 中的文件，满足您对存储数据的管理、分析、检索需求。

智能检索利用数据万象已有的图片、视频、语音、文档等数据处理能力，提取文件的特征或元数据并索引到数据集中，为您提供文件的聚合统计查询、人脸图像检索、图片内容检索等能力。

使用限制

- 智能检索需绑定 COS 存储桶使用，您需要先 [开通对象存储服务](#) 并绑定存储桶。
- 地域限制：仅支持绑定北京、上海地域的存储桶，仅支持对北京、上海地域的 COS 文件建立元数据索引。
- 不同操作中的参数限制，详情请参见 [智能检索 API 接口文档](#)。

 **注意：**

使用智能检索控制台需开通白名单，如有开通需求请 [联系我们](#)。

计费说明

智能检索 MetaInsight 的费用包含算子处理费用、检索费用和元数据管理费用，相关智能检索的计费详细介绍请参见 [智能检索费用](#)。

功能说明

功能	功能描述
管理数据集	数据集是由文件元信息和特征所组成的集合，是存放文件元数据索引的容器。创建数据集后，智能检索 MetaInsight 可以自动解析在对象存储 COS 中文件的基础信息，对图片、视频、音频等媒体文件进行人工智能分析，提取标签、人物等信息并建立元数据索引。您也可以为存储在对象存储 COS 中的文件手动建立元数据索引。
文件元信息查询	通过对文件元信息不同字段的筛选分析，可以查询满足指定条件的文件，并按照指定字段和排序方式列出文件信息。
根据图片检索图片	通过输入图片，用以图搜图的方式可快速在数据集中检索到与输入图片相似的前 N 张图片。
语义检索图	可以根据图片内容输入自然语言进行检索，例如“一片大海”、“一颗饱满的草莓”等图片

片	内容。
人脸检索	您可以使用人脸搜索功能从数据集中搜索与指定图片中人脸最相似的前 N 张图片，适用于门禁闸机、签到考勤、学校宿舍管理等场景。

适用场景

海量文件统计查询

通过智能检索 MetaInsight，您可以从上亿的海量文件中查找符合特定条件的文件或统计出符合特定条件的文件数量。例如查询包含某个关键词的文件，统计某个目录下的文件大小，查询包含某个人物的图片等，您也可以结合您具体的业务情况，通过组合简单查询条件以及聚合操作，做更多场景的数据检索。

数据挖掘

在大数据场景中，采集到的海量图片数据可通过智能检索进行数据预分类，过滤筛选出指定场景的数据后再进行标注训练，可大大提升数据训练效率。

智能相册

使用人脸搜索功能，在图库中搜索与指定人脸最相似的前N张图片，结果按相似度降序排列。在智能相册中可将已经识别的人脸与通讯录进行关联，完成关联后，可在看照片时直接点击照片上的人脸进行打电话、发短信等操作。

电商网站

通过输入商品图片，可以在商品库中准确地找到图片中商品的同款或者相似款，并快速返回对应的商品信息。

使用流程

步骤一：创建数据集

数据集是由文件元信息和特征所组成的集合，是存放文件元数据索引的容器。创建数据集后，智能检索 MetaInsight 可以自动解析在对象存储 COS 中文件的基础信息，并对图片进行智能分析，提取特征建立元数据索引。例如在电商场景中，您创建一个电商图库数据集，通过绑定存储桶与数据集接口将某个存储桶与数据集进行绑定，或通过创建元数据索引接口，为后续产生的电商商品图片建立元数据索引，然后使用图片检索功能，用户可以通过自然语言或相似商品图对电商图库中的图片进行检索。具体操作请参见 [数据集管理控制台指南](#)。

❗ 说明：

在创建数据集时，可以数据集设置检索模板（Template），检索模板定义了数据集创建元数据索引时需要执行的算子操作，不同的算子将提取不同类型的元数据，详情请参见 [检索模板与算子](#)。

步骤二：建立元数据索引

创建数据集后，您可以为存储在对象存储 COS 中的文件建立元数据索引。通过丰富的元数据索引，MetaInsight 可以为您提供强大的数据聚合查询、分析和管理能力。我们提供了两种建立元数据索引的方式：

- **方式一：**绑定 COS 存储桶与数据集，自动建立元数据索引，绑定存储桶后，MetaInsight 会先扫描桶内存量数据，当存量数据索引完成后，会继续扫描桶内新增的文件并建立索引，详情请参见 [数据集管理控制台指南](#)。
- **方式二：**通过接口手动建立元数据索引，详情请参见 [创建元数据索引](#)。

步骤三：数据检索

元数据索引建立完成后，您可以在控制台或通过 API 接口进行基础文件信息查询、以图搜图、以文搜图、人脸搜索等操作，详情请参见 [简单查询](#)、[图像检索](#)、[人脸搜索](#)。

管理文件元数据

数据集管理

创建数据集

最近更新时间：2025-08-29 15:03:52

功能描述

本接口用于创建一个数据集（Dataset），数据集是由文件元数据构成的集合，用于存储和管理元数据索引。

授权说明

通过子账号使用时，需要在 [授权策略](#) 的 action 中添加 `ci:CreateDataset` 权限。数据万象支持的所有操作接口请参见 [CI action](#)。

使用限制

- 每个数据集最多可为1亿个文件建立元数据索引。
- 仅支持北京、上海地域，即请求 Host 中的 `<Region>` 仅支持填写为 `ap-beijing`，`ap-shanghai`。
- COS 加密桶的文件无法被读取并创建元数据索引。

请求

请求示例

```
POST /dataset HTTP/1.1
Host: <AppId>.ci.<Region>.myqcloud.com
Authorization: Auth String
Content-Length: xxx
Content-Type: application/json
Accept: application/json
```

❗ 说明：

Authorization：携带鉴权信息，用于验证请求的合法性的请求头，详情请参见 [请求签名](#) 文档。

请求头

此接口仅使用公共请求头部，详情请参见 [公共请求头部](#) 文档。

请求体

该请求操作的实现需要有以下请求体。

```
{
  "DatasetName": "test",
  "Description": "test",
  "TemplateId": "Official:CosBasicMeta"
}
```

具体的数据内容如下：

参数名称	描述	类型	是否必选
Dataset Name	数据集名称，同一个账户下唯一。命名规则如下： <ul style="list-style-type: none">长度为1 – 32字符只能包含小写英文字母，数字，短划线（-）必须以英文字母和数字开头	String	是
Description	数据集描述信息。长度为1 – 256个英文或中文字符，默认值为空	String	否
TemplateId	<ul style="list-style-type: none">与数据集关联的检索模板，在建立元数据索引时，后端将根据检索模板来决定采集文件的哪些元数据。每个检索模板都包含若干个算子，不同的算子表示不同的处理能力，更多信息请参见 检索模板与算子默认值为空，即不关联检索模板，不进行任何元数据的采集	String	否
Dataset Type	表示数据集的类型，默认为0，表示普通数据集，值为1时表示该数据集为控制台文件列表专用的数据集	Integer	否
Version	数据集版本。basic、standard，默认为 basic	String	否
Volume	当前数据集预估数据量级，MetaInsight 将根据该参数调整数据分布策略以保证查询性能，Version 为 basic 时为50w。Version 为 standard 时，默认为500w，可设置1-10000，单位w。传0采用默认值	Integer	否
Training Mode	训练数据的来源模式。默认为0，表示训练数据来源于指定数据集，值为1时表示训练数据来源于cos某个bucket目录下文件。仅在 Version 为 standard 时生效	Integer	否
Training Dataset	训练数据的数据集名称。仅在 TrainingMode 为0时生效	String	否

Training URI	训练数据的资源路径。仅在 TrainingMode 为1时生效	String	否
SceneType	当 TemplateId=Official:ImageSearch 时生效，表示场景类型。可选值如下： <ul style="list-style-type: none">general（默认值）：通用场景E-commercial：电商场景iDrive：出行场景	String	否

响应

响应头

此接口仅返回公共响应头部，详情请参见 [公共请求头部](#) 文档。

响应体

```
{
  "Dataset": {
    "BindCount": 0,
    "CreateTime": "2023-12-25 15:16:20.692674978 +0800 CST",
    "DatasetName": "test",
    "Description": "test",
    "FileCount": 0,
    "TemplateId": "Official:CosBasicMeta",
    "TotalFileSize": 0,
    "UpdateTime": "2023-12-25 15:16:20.692675128 +0800 CST"
  },
  "RequestId": "NWFjMzQ0MDZfOTBmYTUwXzZkZV8z*****"
}
```

响应包体具体数据内容如下：

参数名称	描述	类型
Dataset	数据集信息。	Container

RequestId	请求 ID。	String
-----------	--------	--------

Dataset 节点内容：

参数名称	描述	类型
BindCount	数据集当前绑定的 COS Bucket 数量	Long
CreateTime	数据集创建时间的时间戳，格式为 RFC3339Nano	String
DatasetName	数据集名称	String
DatasetType	表示数据集的类型，默认为0，表示普通数据集，值为1时表示该数据集为控制台文件列表专用的数据集。	Integer
Description	数据集描述信息	String
FileCount	数据集中当前已建立了元数据索引的文件数量	Long
Region	数据集所在地区	String
SceneType	场景类型。支持 general、E-commercial、iDrive，默认为 general	String
State	训练状态 <ul style="list-style-type: none">Unstart：未开始Running：训练中Success：训练成功Failed：训练失败	String
TaskProgress	训练进度的百分比值。范围为 [0, 100]	Int
TemplateId	检索模板	String
TotalFileSize	数据集中当前已建立了元数据索引的文件的总大小，单位为字节	Long

UpdateTime	<ul style="list-style-type: none">数据集修改时间的时间戳，格式为 RFC3339Nano创建数据集后，如果未更新过数据集，则数据集修改时间的时间戳和数据集创建时间的时间戳相同	String
Version	数据集版本。	String

查询数据集

最近更新时间：2025-11-11 11:53:31

功能描述

查询一个数据集（Dataset）信息。

授权说明

通过子账号使用时，需要在 [授权策略](#) 的 action 中添加 `ci:DescribeDataset` 权限。数据万象支持的所有操作接口请参见 [CI action](#)。

使用限制

使用该接口时，请先确认相关限制条件。详情请参见 [使用限制](#)。

请求

请求示例

```
GET /dataset?datasetname=test&statistics=true HTTP/1.1
Host: <AppId>.ci.<Region>.myqcloud.com
Date: <GMT Date>
Authorization: Auth String
Accept: application/json
```

❗ 说明：

Authorization：携带鉴权信息，用于验证请求的合法性的请求头，详情请参见 [请求签名](#) 文档。

请求头

此接口仅使用公共请求头部，详情请参见 [公共请求头部](#) 文档。

请求参数

参数名称	描述	类型	是否必选
dataset name	数据集名称，同一个账户下唯一。可通过 控制台 或 接口 查询	String	是
statistics	该参数用于指定是否需要立即统计一次数据集中最新的文件数量、大小等数据，参数有效值如下，默认值为 false。	Boolean	否

- **false**：不立即统计，接口返回的文件总大小、数量等数据存在10分钟的延迟
- **true**：立即统计，接口返回的文件总大小、数量等统计数据为实时统计的，统计实时数据需要一定时间，请耐心等待

请求体

该请求无请求体。

响应

响应头

此接口仅返回公共响应头部，详情请参见 [公共响应头部](#) 文档。

响应体

```
{
  "Dataset": {
    "BindCount": 0,
    "CreateTime": "2023-12-25 15:16:20.692674978 +0800 CST",
    "DatasetName": "test",
    "Description": "test",
    "FileCount": 0,
    "TemplateId": "Official:COSEBasicMeta",
    "TotalFileSize": 0,
    "UpdateTime": "2023-12-25 15:16:20.692675128 +0800 CST"
  },
  "RequestId": "NWFjMzQ0MDZfOTBmYTUwXzZkZV8z****"
```

响应包体具体数据内容如下：

参数名称	类型	描述
RequestId	string	请求 ID
Dataset	Container	数据集信息。同创建数据集接口中响应体的 Dataset

更新数据集

最近更新时间：2025-08-22 15:29:12

功能描述

更新一个数据集（Dataset）的配置。

授权说明

通过子账号使用时，需要在 [授权策略](#) 的 action 中添加 `ci:UpdateDataset` 权限。数据万象支持的所有操作接口请参见 [CI action](#)。

限制说明

- 数据集名称无法更改。
- 更新时未填写的参数表示不更新。

请求

请求示例

```
PUT /dataset HTTP/1.1
Host: <AppId>.ci.<Region>.myqcloud.com
Authorization: Auth String
Content-Length: xxx
Content-Type: application/json
Accept: application/json
```

❗ 说明：

Authorization：携带鉴权信息，用于验证请求的合法性的请求头，详情请参见 [请求签名](#) 文档。

请求头

此接口仅使用公共请求头部，详情请参见 [公共请求头部](#) 文档。

请求体

```
{
  "DatasetName": "test",
  "Description": "test",
  "TemplateId": "Official:COSEBasicMeta"
```

```
}
```

请求参数

参数名称	描述	类型	是否必选
DatasetName	数据集名称，同一个账户下唯一。可通过 控制台 或 接口 查询	String	是
Description	数据集描述信息。长度为1~256个英文或中文字符，默认值为空。	String	否
TemplateId	<ul style="list-style-type: none">检索数据集关联的检索模板，在建立元数据索引时，后端将根据检索模板来决定采集文件的哪些元数据。每个检索模板都包含若干个算子，不同的算子表示不同的处理能力，更多信息请参见 检索模板与算子。默认值为空，即不关联检索模板，不进行任何元数据的采集。	String	否

响应

响应头

此接口仅返回公共响应头部，详情请参见 [公共响应头部](#) 文档。

响应体

```
{
  "Dataset": {
    "BindCount": 0,
    "CreateTime": "2023-12-25 15:16:20.692674978 +0800 CST",
    "DatasetName": "test",
    "Description": "test",
    "FileCount": 0,
    "TemplateId": "Official:COSBasicMeta",
    "TotalFileSize": 0,
    "UpdateTime": "2023-12-25 15:16:20.692675128 +0800 CST"
  },
  "RequestId": "NWFjMzQ0MDZfOTBmYTUwXzZkZV8z*****"
}
```

响应包体具体数据内容如下：

参数名称	描述	类型
RequestId	请求 ID。	string
Dataset	数据集信息。同创建数据集接口中响应体的 Dataset	Container

删除数据集

最近更新时间：2025-11-11 11:53:31

功能描述

删除一个数据集（Dataset）。

使用说明

- 删除数据集前，请确保已删除数据集中所有已建立的元数据索引，删除索引文件的方法可参见 [删除元数据索引](#)。
- 删除数据集前，请确保已解绑所有 COS 存储桶和该数据集的绑定关系，解除绑定的方法可参见 [解绑存储桶与数据集](#)。

授权说明

通过子账号使用时，需要在 [授权策略](#) 的 action 中添加 `ci:DeleteDataset` 权限。数据万象支持的所有操作接口请参见 [CI action](#)。

使用限制

使用该接口时，请先确认相关限制条件。详情请参见 [使用限制](#)。

请求

请求示例

```
DELETE /dataset HTTP/1.1
Host: <AppId>.ci.<Region>.myqcloud.com
Authorization: Auth String
Content-Length: xxx
Content-Type: application/json
Accept: application/json
```

❗ 说明：

Authorization：携带鉴权信息，用于验证请求的合法性的请求头，详情请参见 [请求签名](#) 文档。

请求头

此接口仅使用公共请求头部，详情请参见 [公共请求头部](#) 文档。

请求参数

参数名称	描述	类型	是否必选
DatasetName	数据集名称，同一个账户下唯一。可通过 控制台 或 接口 查询	String	是

请求体

```
{
  "DatasetName": "test"
}
```

响应

响应头

此接口仅返回公共响应头部，详情请参见 [公共响应头部](#) 文档。

响应体

```
{
  "Dataset": {
    "BindCount": 0,
    "CreateTime": "2023-12-25 15:16:20.692674978 +0800 CST",
    "DatasetName": "test",
    "Description": "test",
    "FileCount": 0,
    "TemplateId": "Official: COSBasicMeta",
    "TotalFileSize": 0,
    "UpdateTime": "2023-12-25 15:16:20.692675128 +0800 CST"
  },
  "RequestId": "NWFjMzQ0MDZfOTBmYTUwXzZkZV8z*****"
}
```

响应包体具体数据内容如下：

参数名称	描述	类型
RequestId	请求 ID。	string
Dataset	数据集信息。同创建数据集接口中响应体的 Dataset	Contain

		er
--	--	----

列出数据集

最近更新时间：2025-11-11 11:53:31

功能描述

获取数据集（Dataset）列表。

授权说明

通过子账号使用时，需要在 [授权策略](#) 的 action 中添加 `ci:DescribeDatasets` 权限。数据万象支持的所有操作接口请参见 [CI action](#)。

使用限制

使用该接口时，请先确认相关限制条件。详情请参见 [使用限制](#)。

请求

请求示例

```
GET /datasets?maxresults=2&prefix=test&nexttoken=xxxx HTTP/1.1
Host: <AppId>.ci.<Region>.myqcloud.com
Date: <GMT Date>
Authorization: Auth String
Accept: application/json
```



说明：

Authorization: Auth String，详情请参见 [请求签名](#) 文档。

请求头

此接口仅使用公共请求头部，详情请参见 [公共请求头部](#) 文档。

请求参数

参数名称	描述	类型	是否必选
maxresults	本次返回数据集的最大个数，取值范围为0 – 200。不设置此参数或者设置为0时，则默认值为100。	Long	否

nexttoken	翻页标记。当文件总数大于设置的 MaxResults 时，用于翻页的 Token。从 NextToken 开始按字典序返回文件信息列表。填写上次查询返回的值，首次使用时填写为空。	String	否
prefix	数据集名称前缀。	String	否

请求体

该请求无请求体。

响应

响应头

此接口仅返回公共响应头部，详情请参见 [公共响应头部](#) 文档。

响应体

```
{
  "Datasets": [{
    "BindCount": 0,
    "CreateTime": "2023-12-25 15:12:27.139749816 +0800 CST",
    "DatasetName": "test001",
    "Description": "test001",
    "FileCount": 0,
    "TemplateId": "Official:DefaultEmptyId",
    "TotalFileSize": 0,
    "UpdateTime": "2023-12-25 15:12:27.139749953 +0800 CST"
  }, {
    "BindCount": 0,
    "CreateTime": "2023-12-25 15:07:00.331922722 +0800 CST",
    "DatasetName": "test002",
    "Description": "test002",
    "FileCount": 0,
    "TemplateId": "Official:COSEBasicMeta",
    "TotalFileSize": 0,
    "UpdateTime": "2023-12-25 15:41:41.978447866 +0800 CST"
  }],
  "NextToken": "2:1251668577:test",
  "RequestId": "NWFjMzQ0MDZfOTBmYTUwXzZkZV8z*****"
```

```
}

```

响应包体具体数据内容如下：

参数名称	描述	类型
Datasets	数据集信息。同创建数据集接口中响应体的 Dataset	Container Array
NextToken	翻页标记。当任务列表总数大于设置的 MaxResults 时，用于翻页的 Token。符合条件的任务列表未全部返回时，此参数才有值。下一次列出任务列表时将此值作为 NextToken 传入，将后续的任务列表返回。	String
RequestId	请求 ID。	String

绑定存储桶与数据集

最近更新时间：2025-11-17 15:20:22

功能描述

将一个对象存储（COS）存储桶绑定到数据集（Dataset）上。绑定后，将对存储桶中新增的文件进行准实时的增量追踪扫描，依据数据集的检索模板对文件进行处理，抽取文件的元数据进行索引。通过此方式为文件建立元数据索引后，您可以使用智能检索的查询接口进行数据的查询统计。

授权说明

通过子账号使用时，需要在 [授权策略](#) 的 action 中添加 `ci:CreateDatasetBinding` 权限。数据万象支持的所有操作接口请参见 [CI action](#)。

限制说明

- 每个数据集绑定的最大个数为10。
- 目前仅支持绑定北京、上海地域的存储桶到数据集上。
- COS 加密桶的文件无法被读取并创建元数据索引。
- 不支持跨账号的 COS 桶，仅支持绑定本账号下的 COS 桶。

请求

请求示例

```
POST /datasetbinding HTTP/1.1
Host: <AppId>.ci.<Region>.myqcloud.com
Authorization: Auth String
Content-Length: xxx
Content-Type: application/json
Accept: application/json
```

❗ 说明：

Authorization：携带鉴权信息，用于验证请求的合法性的请求头，详情请参见 [请求签名](#) 文档。

请求头

此接口仅使用公共请求头部，详情请参见 [公共请求头部](#) 文档。

请求体

```
{
  "DatasetName": "test",
  "URI": "cos://examplebucket-1250000000/test_data"
}
```

请求参数

参数名称	描述	类型	是否必选
Dataset Name	数据集名称，同一个账户下唯一。可通过 控制台 或 接口 查询	String	是
URI	资源标识字段，表示需要与数据集绑定的资源，当前支持 COS 存储桶及桶内的目录，可添加 prefix 进行目录绑定；字段规则： <code>cos://<BucketName>/prefix</code> ，其中 BucketName 表示 COS 存储桶名称，例如： <code>cos://examplebucket-1250000000</code> ，可直接绑定某存储桶；亦可以添加 prefix，prefix 代表了某个路径，例如： <code>cos://examplebucket-1250000000/test_data</code>	String	是

响应

响应头

此接口仅返回公共响应头部，详情请参见 [公共响应头部](#) 文档。

响应体

```
{
  "Binding": {
    "CreateTime": "2023-12-25 19:19:32.751303843 +0800 CST m=+272347.749940194",
    "DatasetName": "test",
    "State": "Running",
    "URI": "cos://examplebucket-1250000000/test_data",
    "UpdateTime": "2023-12-25 19:19:32.751303991 +0800 CST m=+272347.749940332"
  },
  "RequestId": "NWFjMzQ0MDZfOTBmYTUwXzZkZV8z*****"
}
```

响应包体具体数据内容如下：

参数名称	类型	描述
Binding	Container	绑定信息。
RequestId	String	请求 ID。

Binding 节点内容：

参数名称	类型	描述
Create Time	String	数据集和 COS Bucket 绑定关系创建时间的时间戳，格式为 RFC3339Nano。
Dataset Name	String	数据集名称。
State	String	数据集和 COS Bucket 绑定关系的状态。取值范围如下： Running：绑定关系运行中。
URI	String	资源标识字段，表示需要与数据集绑定的资源，当前支持 COS 存储桶及桶内的目录，可添加prefix进行目录绑定；字段规则： <code>cos://<BucketName>/prefix</code> ，其中 BucketName 表示 COS 存储桶名称，例如： <code>cos://examplebucket-1250000000</code> ，可直接绑定某存储桶；亦可以添加prefix，prefix代表了某个路径，例如： <code>cos://examplebucket-1250000000/test_data</code>
Update Time	String	<ul style="list-style-type: none">数据集和 COS Bucket 的绑定关系修改时间的时间戳，格式为 RFC3339Nano。创建绑定关系后，如果未暂停或者未重启过绑定关系，则绑定关系修改时间的时间戳和绑定关系创建时间的时间戳相同。
StockState	String	当前绑定的存储桶对应的存量索引的状态：有效值如下： <ul style="list-style-type: none">NoIndexing：未进行存量建立索引Indexing：存量索引建立中Success：存量索引已建立完成

解绑存储桶与数据集

最近更新时间：2025-11-11 11:53:31

功能描述

将一个与数据集绑定了的对象存储（COS）存储桶进行解绑，解绑后，存储桶内新增的文件不会再自动建立元数据索引到数据集。

授权说明

通过子账号使用时，需要在 [授权策略](#) 的 action 中添加 `ci:DeleteDatasetBinding` 权限。数据万象支持的所有操作接口请参见 [CI action](#)。

使用限制

使用该接口时，请先确认相关限制条件。详情请参见 [使用限制](#)。

请求

请求示例

```
DELETE /datasetbinding HTTP/1.1
Host: <AppId>.ci.<Region>.myqcloud.com
Authorization: Auth String
Content-Length: xxx
Content-Type: application/json
Accept: application/json
```

❗ 说明：

Authorization：携带鉴权信息，用于验证请求的合法性的请求头，详情请参见 [请求签名](#) 文档。

请求头

此接口仅使用公共请求头部，详情请参见 [公共请求头部](#) 文档。

请求参数

参数名称	描述	类型	是否必选
Dataset Name	数据集名称，同一个账户下唯一。可通过 控制台 或 接口 查询	String	是

URI	资源标识字段，表示需要与数据集绑定的资源，当前支持 COS 存储桶及桶内的目录，可添加 prefix 进行目录绑定；字段规则： <code>cos://<BucketName>/prefix</code> ，其中 BucketName 表示 COS 存储桶名称，例如： <code>cos://examplebucket-1250000000</code> ，可直接绑定某存储桶；亦可以添加 prefix，prefix 代表了某个路径，例如： <code>cos://examplebucket-1250000000/test_data</code>	String	是
-----	---	--------	---

请求体

```
{
  "DatasetName": "test",
  "URI": "cos://examplebucket-1250000000"
}
```

响应

响应头

此接口仅返回公共响应头部，详情请参见 [公共响应头部](#) 文档。

响应体

```
{
  "RequestId": "NjU4OTZhOGZfNGQ2ODk0MGJfNWEyMF81"
}
```

响应包体具体数据内容如下：

参数名称	类型	描述
RequestId	String	请求 ID。

查询数据集与存储桶的绑定关系

最近更新时间：2025-11-11 11:53:31

功能描述

查询数据集和某个指定对象存储（COS）Bucket 的绑定关系。

授权说明

通过子账号使用时，需要在 [授权策略](#) 的 action 中添加 `ci:DescribeDatasetBinding` 权限。数据万象支持的所有操作接口请参见 [CI action](#)。

使用限制

使用该接口时，请先确认相关限制条件。详情请参见 [使用限制](#)。

请求

请求示例

```
GET /datasetbinding?datasetname=test&uri=cos%3A%2F%2Fexample-1250000000
HTTP/1.1
Host: <AppId>.ci.<Region>.myqcloud.com
Date: <GMT Date>
Authorization: Auth String
Accept: application/json
```

❗ 说明：

Authorization：携带鉴权信息，用于验证请求的合法性的请求头，详情请参见 [请求签名](#) 文档。

请求头

此接口仅使用公共请求头部，详情请参见 [公共请求头部](#) 文档。

请求参数

参数名称	描述	类型	是否必选
datasetname	数据集名称，同一个账户下唯一。可通过 控制台 或 接口 查询。	String	是

uri	资源标识字段，表示需要查询与数据集绑定关系的资源，当前仅支持 COS 存储桶，字段规则： <code>cos://<BucketName></code> ，其中 BucketName 表示 COS 存储桶名称，例如（需要进行 URL Encode）： <code>cos%3A%2F%2Fexample-1250000000</code> 。	String	是
-----	---	--------	---

请求体

该请求的请求体为空。

响应

响应头

此接口仅返回公共响应头部，详情请参见 [公共响应头部](#) 文档。

响应体

```
{
  "Binding": {
    "CreateTime": "2023-12-26 17:28:33.617269703 +0800 CST",
    "DatasetName": "test",
    "State": "Running",
    "StockState": "Indexing",
    "URI": "cos://examplebucket2-1250000000",
    "UpdateTime": "2023-12-26 17:28:33.617269858 +0800 CST"
  },
  "RequestId": "NWFjMzQ0MDZfOTBmYTUwXzZkZV8z*****"
}
```

响应包体具体数据内容如下：

参数名称	类型	描述
Binding	Container	数据集和 COS Bucket 绑定关系信息。
RequestId	String	请求 ID。

Binding 节点内容：

参数名称	类型	描述
CreateTime	String	数据集和 COS Bucket 绑定关系创建时间的的时间戳，格式为 RFC3339Nano。

DatasetName	String	数据集名称。
State	String	数据集和 COS Bucket 绑定关系的状态。取值范围如下： Running：绑定关系运行中。
StockState	String	当前绑定的存储桶对应的存量索引的状态：有效值如下： <ul style="list-style-type: none">NoIndexing：未进行存量建立索引Indexing：存量索引建立中Success：存量索引已建立完成
URI	String	资源标识字段，表示需要查询与数据集绑定关系的资源，当前仅支持 COS 存储桶，字段规则： <code>cos://<BucketName></code> ，其中 BucketName 表示 COS 存储桶名称，例如： <code>cos://examplebucket-1250000000</code> 。
UpdateTime	String	<ul style="list-style-type: none">数据集和 COS Bucket 的绑定关系修改时间的时间戳，格式为 RFC3339Nano。创建绑定关系后，如果未暂停或者未重启过绑定关系，则绑定关系修改时间的时间戳和绑定关系创建时间的时间戳相同。

查询绑定关系列表

最近更新时间：2025-11-11 11:53:31

功能描述

查询数据集和对象存储（COS）Bucket 绑定关系列表。

授权说明

通过子账号使用时，需要在 [授权策略](#) 的 action 中添加 `ci:DescribeDatasetBindings` 权限。数据万象支持的所有操作接口请参见 [CI action](#)。

使用限制

使用该接口时，请先确认相关限制条件。详情请参见 [使用限制](#)。

请求

请求示例

```
GET /datasetbindings?datasetname=test&limit=100 HTTP/1.1
Host: <AppId>.ci.<Region>.myqcloud.com
Date: <GMT Date>
Authorization: Auth String
Accept: application/json
```

❗ 说明：

Authorization：携带鉴权信息，用于验证请求的合法性的请求头，详情请参见 [请求签名](#) 文档。

请求头

此接口仅使用公共请求头部，详情请参见 [公共请求头部](#) 文档。

请求参数

具体参数如下：

参数名称	描述	类型	是否必选
datasetname	数据集名称，同一个账户下唯一。可通过 控制台 或 接口 查询	String	是

maxresults	返回绑定关系的最大个数，取值范围为0~200。不设置此参数或者设置为0时，则默认值为100。	Integer	否
nexttoken	当绑定关系总数大于设置的 MaxResults 时，用于翻页的 token。从 NextToken 开始按字典序返回绑定关系信息列表。第一次调用此接口时，设置为空。	String	否

请求体

该请求无请求体。

响应

响应头

此接口仅返回公共响应头部，详情请参见 [公共响应头部](#) 文档。

响应体

```
{
  "Bindings": [{
    "CreateTime": "2023-12-25 19:18:19.809729971 +0800 CST",
    "DatasetName": "test",
    "State": "Running",
    "URI": "cos://examplebucket1-1250000000",
    "UpdateTime": "2023-12-25 19:18:19.809730114 +0800 CST"
  }, {
    "CreateTime": "2023-12-25 19:18:19.809729971 +0800 CST",
    "DatasetName": "test",
    "State": "Running",
    "URI": "cos://examplebucket2-1250000000",
    "UpdateTime": "2023-12-25 19:18:19.809730114 +0800 CST"
  }],
  "NextToken": "2:1250000000:test",
  "RequestId": "NWFjMzQ0MDZfOTBmYTUwXzZkZV8z*****"
}
```

响应包体具体数据内容如下：

参数名称	类型	描述
Binding	Contain	数据集和 COS Bucket 绑定关系信息的列表。

s	er Array	
NextToken	String	当绑定关系总数大于设置的 MaxResults 时，用于翻页的 token。下一次列出绑定关系信息时以此值为 NextToken，将未返回的结果返回。当绑定关系未全部返回时，此参数才有值。
RequestId	String	请求 ID。

Bindings 节点内容：

参数名称	类型	描述
CreateTime	String	数据集和 COS Bucket 绑定关系创建时间的时间戳，格式为 RFC3339Nano。
DatasetName	String	数据集名称。
State	String	数据集和 COS Bucket 绑定关系的状态。取值范围如下： Running：绑定关系运行中。
URI	String	资源标识字段，表示需要与数据集绑定的资源，当前仅支持 COS 存储桶，字段规则： <code>cos://<BucketName></code> ，其中BucketName 表示 COS 存储桶名称，例如： <code>cos://examplebucket-1250000000</code> 。
UpdateTime	String	数据集和 COS Bucket 的绑定关系修改时间的时间戳，格式为 RFC3339Nano。 创建绑定关系后，如果未暂停或者未重启过绑定关系，则绑定关系修改时间的时间戳和绑定关系创建时间的时间戳相同。
StockState	String	当前绑定的存储桶对应的存量索引的状态：有效值如下： <ul style="list-style-type: none">NoIndexing：未进行存量建立索引Indexing：存量索引建立中Success：存量索引已建立完成

检索模板与算子

最近更新时间：2024-06-12 17:46:52

简介

在创建数据集时，可以为数据集设置检索模板（TemplateId），检索模板可以包含一个或多个算子，算子决定了数据集中需要收集哪些元数据。在为文件建立元数据索引时，后端将根据检索模板中的算子，采集文件的基础信息、内容特征等变为元数据索引存储到数据集中。


检索模板

目前仅提供官方预设检索模板，如下所示：

检索模板 (TemplateId)	说明
Official:CosBasicMeta	COS 基础信息模板。在建立元数据索引时，执行 COS 基础元信息算子，会采集 COS 文件的名称、大小、存储类型等基础信息，适用于对 COS 中的文件进行目录统计、关键词模糊检索等。
Official:FaceSearch	人脸搜索模板。在建立元数据索引时，执行 COS 基础元信息算子和人脸信息算子，会采集图片中的人脸信息，适用于人物相册等场景。
Official:ImageSearch	图像搜索模板。在建立元数据索引时，执行 COS 基础元信息算子和图片特征算子，会采集图片中的内容特征，适用于智能相册、图片分类等场景。

算子

算子是采集元数据的最小处理单元，不同算子代表不同的数据处理能力，通过数据处理采集不同的文件元数据。

 **注意：**
使用算子可能产生相应的计费费用，详情请查看 [智能检索费用](#)。

算子	说明
COS基础元信息算子	获取 COS 文件的基础元信息，如文件名、文件类型、文件大小、文件 HASH、文件修改时间、ETag、标签等数据，具体信息请参见 字段和操作符的支持列表 。
人脸信息算子	通过人脸识别能力，获取图片中的人脸信息。
图片特征算子	通过图片分析能力，获取图片的内容特征。

检索模板与算子的对应关系

TemplateId	COS基础信息算子	人脸信息算子	图片特征算子
Official:COSBasicMeta	包含	不包含	不包含
Official:FaceSearch	包含	包含	不包含
Official:ImageSearch	包含	不包含	包含

数据索引

数据索引概述

最近更新时间：2024-12-03 10:07:11

自动建立数据索引，利用文件的元数据和向量语义作为查询条件，快速查找 COS 中的图片、视频、文档、音频文件。

COS 数据索引的优势

- 直连 COS：直接使用 COS 数据构建的数据索引，无需搬迁数据和自建检索系统。
- 多模态：支持多种索引类型，包括 OSS 元数据、媒体元数据、自定义元数据、向量语义；提供近百种检索条件。
- 高性能：实现秒级索引和聚合，支持亿级文件索引的数据集，满足大规模数据处理需求。

支持的数据检索方式

COS 支持标量检索和向量检索两种检索方式：

	标量检索	向量检索
定义	基于 COS 元数据、对象 ETag、对象标签等元数据属性进行匹配查询	通过将图片文件的信息表示为向量，并利用这些向量进行语义相似性的比较和检索，实现根据文件内容进行语义查询
使用场景	文件查询、资产盘点	图库语义查询
查询条件示例	查询2024年11月1日以后上传的、私有的、标准存储的文件	查询和“番茄”相关的图片
返回结果示例	<div>返回2024年11月1日以后上传的、标准存储的文件列表</div> <div></div>	<div>返回和“番茄”相关的图片文件列表</div> <div></div>

选择数据检索方式

根据所需的检索条件与适用场景，参考下表选择标量检索或向量检索：

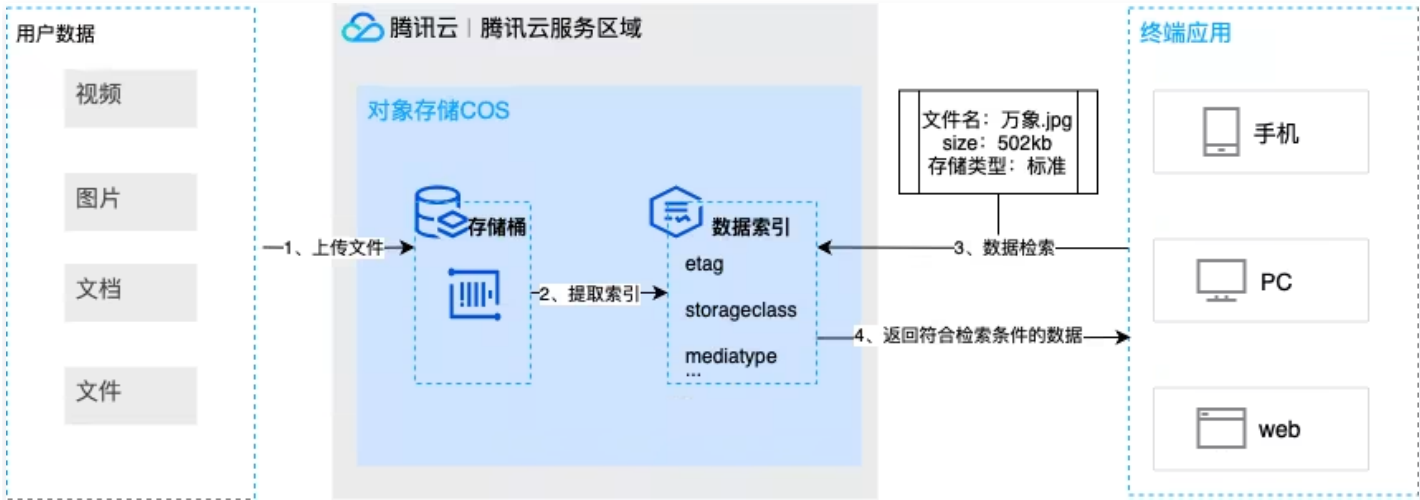
检索条件	适用场景	检索方式
------	------	------

COS元数据	资产盘点	标量检索
对象标签和对象ETag	数据分类统计	标量检索
自定义元数据	数据分类统计	标量检索
媒体信息	媒资管理	标量检索
向量语义（自然语言或图片）	电商商品推荐、网盘相册	向量检索

说明：
关于标量检索支持的所有元数据字段，请参见 [字段和操作符的支持列表](#)。

开始使用数据检索

使用标量或向量检索 COS 数据的流程如下：



- 用户上传图片、视频、文档、音频等文件到 COS Bucket。
- 具备 COS 管理权限的用户为 Bucket 开启智能检索后将自动提取 COS 桶内的数据索引。
- 自动建立包含 COS 元数据、对象 ETag、对象标签的数据索引。
- 终端应用调用 [简单查询](#) API 基于元数据属性进行查询。
- COS 返回满足查询条件的数据。

说明：
COS 数据检索可通过控制台进行操作，详情请参见 [标量检索](#)、[向量检索](#)。

开始进行数据检索

进行标量检索和向量检索的详细步骤请参见：

- [标量检索](#)

- [向量检索](#)

索引操作

创建元数据索引

最近更新时间：2025-11-17 15:20:22

功能描述

在数据集中建立一个 COS 文件的元数据索引，会根据数据集的检索模板提取元数据去建立索引，支持额外添加自定义的元数据。

授权说明

通过子账号使用时，需要在 [授权策略](#) 的 action 中添加 `ci:CreateFileMetaIndex` 权限。数据万象支持的所有操作接口请参见 [CI action](#)。

使用说明

- 文件元数据建立索引数据处理支持列表请参见 [检索模板与算子](#)。
- 文件元数据建立索引有文件总数量限制或加密文件的限制，请参见 [创建数据集](#) 使用参数限制说明。
- 文件元数据建立索引后，您可以通过查询接口进行数据检索。
- 该接口为异步操作，在您提交请求后，需要一定时间对您提交的文件进行处理，处理完成后才会将元数据存储到数据集内。该过程视数据集所用的检索模板不同、文件内容的不同，需要数秒到数分钟甚至更久。

使用限制

使用该接口时，请先确认相关限制条件。详情请参见 [使用限制](#)。

请求

请求示例

```
POST /filemeta HTTP/1.1
Host: <AppId>.ci.<Region>.myqcloud.com
Authorization: Auth String
Content-Length: xxx
Content-Type: application/json
Accept: application/json
```



说明：

Authorization: Auth String，详情请参见 [请求签名](#) 文档。

请求头

此接口仅使用公共请求头部，详情请参见 [公共请求头部](#) 文档。

请求体

该请求操作的实现需要有以下请求体。

```
{
  "DatasetName": "test001",
  "File": {
    "URI": "cos://examplebucket-1250000000/test.jpg",
    "CustomId": "001",
    "CustomLabels": {
      "age": "18",
      "level": "18"
    },
    "MediaType": "image",
    "ContentType": "image/jpeg",
    "MaxFaceNum": 20,
    "Persons": [{
      "PersonId": "xxxxxx"
    }]
  }
}
```

具体的参数如下：

参数名称	描述	类型	是否必选
Dataset Name	数据集名称，同一个账户下唯一。可通过 控制台 或 接口 查询。	String	是
File	待建立元数据索引的文件信息及自定义元数据内容。	Container	是

File 节点内容：

参数名称	描述	类型	是否必选
URI	资源标识字段，表示需要建立索引的文件地址，当前仅支持 COS 上的文件，字段规则： <code>cos://<BucketName>/<ObjectKey></code> ，其中 BucketName 表示 COS 存储桶名称，	String	是

	<p>ObjectKey 表示文件完整路径，例如：<code>cos://examplebucket-1250000000/test1/img.jpg</code>。</p> <p>注意：</p> <ul style="list-style-type: none">仅支持本账号内的 COS 文件。不支持 HTTP 开头的地址。		
CustomId	<p>自定义 ID。该文件索引到数据集后，作为该行元数据的属性存储，用于和您的业务系统进行关联、对应。您可以根据业务需求传入该值，例如将某个 URI 关联到您系统内的某个 ID。推荐传入全局唯一的值。在查询时，该字段支持前缀查询和排序，详情请见 字段和操作符的支持列表。</p>	String	否
CustomLabels	<p>自定义标签。您可以根据业务需要自定义添加标签键值对信息，用于在查询时可以据此为筛选项进行检索，详情请见 字段和操作符的支持列表。</p>	Object	否
MediaType	<p>可选项，文件媒体类型，枚举值：</p> <ul style="list-style-type: none">image：图片。document：文档。archive：压缩包。video：视频。audio：音频。other：其他。	String	否
ContentType	<p>可选项，文件内容类型（MIME Type），如 image/jpeg。</p>	String	否
MaxFaceNum	<p>输入图片中检索的人脸数量，默认值为20，最大值为20。（仅当数据集的检索模板 ID 为 Official:FaceSearch 有效）。</p>	Integer	否
Persons	<p>自定义人物属性(仅当数据集的检索模板 ID 为 Official:FaceSearch 有效)。</p>	Container Array	否

Persons 节点内容：

参数名称	描述	类型	是否必选
PersonId	自定义人物 ID。	String	否

响应

响应头

此接口仅返回公共响应头部，详情请参见 [公共响应头部](#) 文档。

响应体

```
{
  "EventId": "wie3580c6ca88dd83382cf52540067****",
  "RequestId": "NWFjMzQ0MDZfOTBmYTUwXzZkZV8z****"
}
```

响应包体具体数据内容如下：

参数名称	类型	描述
RequestId	String	请求 ID。
EventId	String	创建元数据索引的任务 ID。

更新元数据索引

最近更新时间：2025-11-11 11:53:31

功能描述

更新数据集内已索引的文件的部分元数据。

并非所有的元数据都允许您自定义更新，在您发起更新请求时需要填写数据集，会根据该数据集的检索模板重新提取元数据并更新已存在的元数据索引，此外您也可以更新部分自定义的元数据索引，如 CustomLabels、CustomId 等字段，具体请参考 [请求参数说明](#)。

授权说明

通过子账号使用时，需要在 [授权策略](#) 的 action 中添加 `ci:UpdateFileMetaIndex` 权限。数据万象支持的所有操作接口请参见 [CI action](#)。

使用限制

使用该接口时，请先确认相关限制条件。详情请参见 [使用限制](#)。

请求

请求示例

```
PUT /filemeta HTTP/1.1
Host: <AppId>.ci.<Region>.myqcloud.com
Authorization: Auth String
Content-Length: xxx
Content-Type: application/json
Accept: application/json
```

❗ 说明：

Authorization：携带鉴权信息，用于验证请求合法性的请求头，详情请参见 [请求签名](#) 文档。

请求头

此接口仅使用公共请求头部，详情请参见 [公共请求头部](#) 文档。

请求体

```
{
  "DatasetName": "test001",
```

```
"File": {
  "URI": "cos://examplebucket-1250000000/test1/img.jpg",
  "CustomId": "001",
  "CustomLabels": {
    "age": "18",
    "level": "18"
  },
  "MediaType": "image",
  "ContentType": "image/jpeg"
}
```

具体的参数如下：

参数名称	描述	类型	是否必选
Dataset Name	数据集名称，同一个账户下唯一。可通过 控制台 或 接口 查询	String	是
File	用于建立索引的文件信息。	Container	是

File 节点内容：

参数名称	描述	类型	是否必选
URI	资源标识字段，表示需要建立索引的文件地址，当前仅支持 COS 上的文件，字段规则： <code>cos://<BucketName>/<ObjectKey></code> ，其中 BucketName 表示 COS 存储桶名称，ObjectKey 表示文件完整路径，例如： <code>cos://examplebucket-1250000000/test1/img.jpg</code> 。 注意： <ul style="list-style-type: none">仅支持本账号内的 COS 文件。不支持 HTTP 开头的地址。	String	是
CustomId	自定义 ID。该文件索引到数据集后，作为该行元数据的属性存储，用于和您的业务系统进行关联、对应。您可以根据业务需求传入该值，例如将某个 URI 关联到您系统内的某个 ID。推荐传入全局唯一的值。在查询时，该字段支持前缀查询和排序，详情请见 字段和操作符的支持列表 。	String	否

Custom Labels	自定义标签。您可以根据业务需要自定义添加标签键值对信息，用于在查询时可以据此为筛选项进行检索，详情请见 字段和操作符的支持列表 。	Object	否
Key	自定义标签键。详情请见 字段和操作符的支持列表 。	String	否
Value	自定义标签值。详情请见 字段和操作符的支持列表 。	String	否
MediaType	可选项，文件媒体类型，枚举值： <ul style="list-style-type: none">image：图片。document：文档。archive：压缩包。video：视频。audio：音频。other：其他。	String	否
ContentType	可选项，文件内容类型（MIME Type），如 <code>image/jpeg</code> 。	String	否

响应

响应头

此接口仅返回公共响应头部，详情请参见 [公共响应头部](#) 文档。

响应体

```
{
  "EventId": "wie3580c6ca88dd83382cf52540067****",
  "RequestId": "NWFjMzQ0MDZfOTBmYTUwXzZkZV8z****"
}
```

响应包体具体数据内容如下：

参数名称	类型	描述
EventId	String	创建元数据索引的任务 ID。
RequestId	String	请求 ID。

查询元数据索引

最近更新时间：2025-11-11 11:53:31

功能描述

获取数据集内已完成索引的一个文件的元数据。

授权说明

通过子账号使用时，需要在 [授权策略](#) 的 action 中添加 `ci:DescribeFileMetaIndex` 权限。数据万象支持的所有操作接口请参见 [CI action](#)。

使用限制

使用该接口时，请先确认相关限制条件。详情请参见 [使用限制](#)。

请求

请求示例

```
GET /filemeta?datasetname=test&uri=cos%3A%2F%2Fexamplebucket-12500000000%2Ftest.jpg HTTP/1.1
Host: <AppId>.ci.<Region>.myqcloud.com
Date: <GMT Date>
Authorization: Auth String
Accept: application/json
```

❗ 说明：

- Authorization: Auth String（详情请参见 [请求签名](#) 文档）。
- 通过子账号使用时，需要授予相关的权限，详情请参见 [授权粒度详情](#) 文档。

请求头

此接口仅使用公共请求头部，详情请参见 [公共请求头部](#) 文档。

请求参数

参数名称	描述	类型	是否必选
------	----	----	------

dataset name	数据集名称，同一个账户下唯一。可通过 控制台 或 接口 查询	String	是
uri	<p>资源标识字段，表示需要建立索引的文件地址，当前仅支持 COS 上的文件，字段规则：<code>cos://<BucketName>/<ObjectKey></code>，其中BucketName表示 COS 存储桶名称，ObjectKey 表示文件完整路径，例如：<code>cos://examplebucket-1250000000/test1/img.jpg</code>。</p> <p>注意：</p> <ul style="list-style-type: none">● 仅支持本账号内的 COS 文件● 不支持 HTTP 开头的地址● 需 UriEncode	String	是

请求体

该请求无请求体。

响应

响应头

此接口仅返回公共响应头部，详情请参见 [公共响应头部](#) 文档。

响应体

```
{
  "Files": [{
    "COSCRC64": "1111122467985959934",
    "COSStorageClass": "STANDARD",
    "ContentType": "image/jpeg",
    "CreateTime": "2023-12-26T11:43:18.92077679+08:00",
    "CustomId": "001",
    "CustomLabels": {
      "age": "18",
      "level": "18"
    },
    "DatasetName": "test",
    "ETag": "\"21aeb337eb76df70dcb1e2618598****\"",
    "FileModifiedTime": "2023-06-07T07:18:04Z",
    "Filename": "examplebucket-1250000000/test.jpg",
    "MediaType": "image",
    "ObjectACL": "default",
```

```
    "Size": "10000",
    "URI": "cos://examplebucket-1250000000/test.jpg",
    "UpdateTime": "2023-12-26T11:45:02.445509346+08:00"
  },
  "RequestId": "NWFjMzQ0MDZfOTBmYTUwXzZkZV8z*****"
}
```

响应包体具体数据内容如下：

参数名称	类型	描述
Files	Container Array	文件元数据的结构体。实际返回的数据可能并不包含该结构体的所有属性，这和您索引该文件时选用的 workflow 模板配置以及文件本身的内容有关。
RequestId	String	请求 ID。

Files 节点内容：

参数名称	类型	描述
COSCRC64	String	文件 CRC64 值。
COSStorageClass	String	文件存储空间类型。
ContentType	String	文件内容类型（MIME Type）。
CreateTime	String	元数据创建时间的时间戳，格式为 RFC3339Nano。
CustomId	String	该文件的自定义 ID。该文件索引到数据集后，作为该行元数据的属性存储，用于和您的业务系统进行关联、对应。您可以根据业务需求传入该值，例如将某个 URI 关联到您系统内的某个 ID。推荐传入全局唯一的值。
CustomLabels	Container	文件自定义标签列表。储存您业务自定义的键名、键值对信息，用于在查询时可以据此为筛选项进行检索。
DatasetName	String	数据集名称。

ETag	String	Object 生成时会创建相应的 ETag，ETag 用于标识一个 Object 的内容。
FileModifiedTime	String	文件最近一次修改时间的时间戳，格式为 RFC3339Nano。
Filename	String	文件路径。
MediaType	String	文件媒体类型。枚举值： <ul style="list-style-type: none"> • image：图片。 • other：其他。 • document：文档。 • archive：压缩包。 • audio：音频。 • video：视频。
ObjectACL	String	对象的访问控制列表。
Size	Int	文件大小，单位为字节。
URI	String	资源标识字段，表示需要建立索引的文件地址。
UpdateTime	String	<ul style="list-style-type: none"> • 元数据修改时间的时间戳，格式为 RFC3339Nano。 • 创建元数据后，如果未更新过元数据，则元数据修改时间的时间戳和元数据创建时间的时间戳相同。

删除元数据索引

最近更新时间：2025-11-11 11:53:31

功能描述

从数据集内删除一个文件的元信息。

⚠ 注意：

无论该文件的元信息是否存在于数据集内，均会返回删除成功。

授权说明

通过子账号使用时，需要在 [授权策略](#) 的 action 中添加 `ci:DeleteFileMetaIndex` 权限。数据万象支持的所有操作接口请参见 [CI action](#)。

使用限制

使用该接口时，请先确认相关限制条件。详情请参见 [使用限制](#)。

请求

请求示例

```
DELETE /filemeta HTTP/1.1
Host: <AppId>.ci.<Region>.myqcloud.com
Authorization: Auth String
Content-Length: xxx
Content-Type: application/json
Accept: application/json
```

ⓘ 说明：

- Authorization: Auth String（详情请参见 [请求签名](#) 文档）。
- 通过子账号使用时，需要授予相关的权限，详情请参见 [授权粒度详情](#) 文档。

请求头

此接口仅使用公共请求头部，详情请参见 [公共请求头部](#) 文档。

请求体

```
{
  "DatasetName": "test",
  "URI": "cos://examplebucket-1250000000/test.jpg"
}
```

响应

请求参数

参数名称	描述	类型	是否必选
DatasetName	数据集名称，同一个账户下唯一。可通过 控制台 或 接口 查询	String	是
URI	资源标识字段，表示需要删除索引的文件地址，当前仅支持 COS 上的文件，字段规则： <code>cos://<BucketName>/<ObjectKey></code> ，其中 <code>BucketName</code> 表示 COS 存储桶名称， <code>ObjectKey</code> 表示文件完整路径，例如： <code>cos://examplebucket-1250000000/test1/img.jpg</code> 。 注意： <ul style="list-style-type: none">仅支持本账号内的 COS 文件。不支持 HTTP 开头的地址。	String	是

响应

响应头

此接口仅返回公共响应头部，详情请参见 [公共响应头部](#) 文档。

响应体

```
{
  "RequestId": "NWFjMzQ0MDZfOTBmYTUwXzZkZV8z*****"
}
```

响应包体具体数据内容如下：

参数名称	类型	描述
RequestId	String	请求 ID。

标量检索

标量检索概述

最近更新时间：2025-09-02 15:52:52

标量检索是基于 COS 对象元数据的索引功能，允许用户自定义检索条件，快速筛选并获取目标 Object。可以帮助用户更好地管理与了解数据结构，方便后续查询、统计和管理 Object。

使用场景

资产盘点

用户可以通过标量检索对企业数据资产快速统计清点，例如根据媒体类型进行资产数量统计，无需数据库操作流程介入，提升盘点效率与安全性。

数据审计场景

使用标量检索，用户可以快速定位文件，以满足数据审计或监管需求。例如，在金融服务行业，您可以通过自定义标签、访问权限等元数据进行筛选，从而找到具有特定敏感性级别或特定权限的文件，以提升数据审计的效率。

企业备份与归档

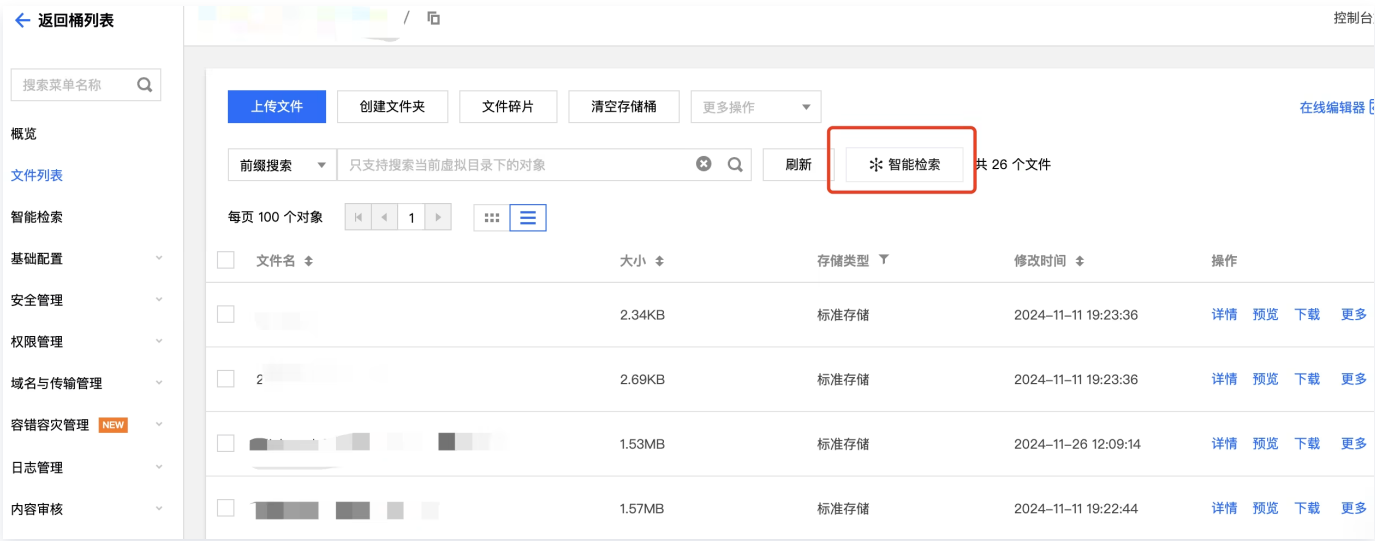
企业在进行数据备份和归档时，可以利用标量检索，例如根据文件的创建时间、存储类型或自定义标签等元数据，迅速检索到特定日期或类型的文件，从而快速恢复历史数据或归档记录。

❗ 说明：

目前该能力仅在北京、上海地域开放。

操作步骤

1. 登录 [对象存储控制台](#)。
2. 在左侧导航栏中，单击**存储桶列表**，进入存储桶列表页面。
3. 找到对象所在的存储桶，单击其**存储桶名称**，进入存储桶管理页面。
4. 在左侧导航栏中，选择**文件列表**，进入文件列表页面。
5. 单击右侧**智能检索**。



6. 在弹窗中确认如下配置。



- **创建数据集：**智能检索将自动为用户创建默认数据集。如果当前操作的存储桶之前使用过智能检索服务，则可能已经存在直接可用的默认的数据集，此时只需要绑定数据集即可。
- **绑定存储桶：**默认绑定当前操作文件列表所在的存储桶。
- **数据索引规则：**可选择仅索引开启智能检索后新上传的文件，也可选择全量数据，对桶内已有的存量数据进行索引。

7. 单击**立即创建**，进入当前存储桶智能检索操作页面。



8. 可在基础信息检索处开始标量检索操作，您可使用文件名称、存储类型、文件大小等信息进行文件查询与聚合统计。

说明：
标量检索字段说明请参见 [字段和操作符的支持列表](#)。

简单查询

最近更新时间：2025-08-29 15:03:52

功能描述

可以根据已提取的文件元数据（包含文件名、标签、路径、自定义标签、文本等字段）查询和统计数据集内文件，支持逻辑关系表达方式。

授权说明

通过子账号使用时，需要在 [授权策略](#) 的 action 中添加 `ci:DatasetSimpleQuery` 权限。数据万象支持的所有操作接口请参见 [CI action](#)。

限制说明

- 每次查询最多返回100个文件信息。
- 每次查询最多返回2000条聚合统计信息。
- 子查询条件最大支持设置100个。
- 子查询嵌套深度最多支持5层。

请求

请求示例

```
POST /datasetquery/simple HTTP/1.1
Host: <AppId>.ci.<Region>.myqcloud.com
Authorization: <Auth String>
Content-Length: <length>
Content-Type: application/json
Accept: application/json
```

❗ 说明：

Authorization: Auth String，详情请参见 [请求签名](#) 文档。

请求头

此接口仅使用公共请求头部，详情请参见 [公共请求头部](#) 文档。

请求体

```
{
  "DatasetName": "test",
  "Query": {
    "Operation": "and",
    "SubQueries": [{
      "Field": "ContentType",
      "Value": "image/jpeg",
      "Operation": "eq"
    },
    {
      "Field": "Size",
      "Value": "1000",
      "Operation": "gt"
    }
  ]
},
  "Sort": "CustomId",
  "Order": "desc",
  "MaxResults": 100
}
```

请求参数

参数名称	描述	类型	是否必选
Dataset Name	数据集名称，同一个账户下唯一。可通过 控制台 或 接口 查询。	String	是
Query	简单查询参数条件，可自嵌套。	Container	否
Sort	排序字段列表。请参考 字段和操作符的支持列表 。 <ul style="list-style-type: none"> 多个排序字段可使用半角逗号（,）分隔，例如：Size,Filename。 最多可设置5个排序字段。 排序字段顺序即为排序优先级顺序。 	String	否
Order	排序字段的排序方式。取值如下： <ul style="list-style-type: none"> asc：升序； desc（默认）：降序。 	String	否

	<p>规则如下：</p> <ul style="list-style-type: none"> 多个排序方式可使用半角逗号（,）分隔，例如：asc,desc。 排序方式不可多于排序字段，即参数 Order 的元素数量需小于等于参数 Sort 的元素数量。例如 Sort 取值为 Size,Filename 时，Order 可取值为 asc,desc 或 asc。 排序方式少于排序字段时，未排序的字段默认取值 asc。例如 Sort 取值为 Size,Filename，Order 取值为 asc 时，Filename 默认排序方式为 asc，即升序排列。 		
MaxResults	<p>返回文件元数据的最大个数。</p> <ul style="list-style-type: none"> 取值范围为0 – 200。 使用聚合参数时，该值表示返回分组的最大个数，取值范围为0 – 2000。 不设置此参数或者设置为0时，则取默认值100。 	Integer	否
NextToken	<p>当绑定关系总数大于设置的 MaxResults 时，用于翻页的 token。</p> <ul style="list-style-type: none"> 从 NextToken 开始按字典序返回绑定关系信息列表。 第一次调用此接口时，设置为空。 	String	否
Aggregations	<p>聚合字段信息列表。</p> <p>当您使用聚合查询时，仅返回聚合结果，不再返回匹配到的元信息列表。</p>	Container Array	否
WithFields	<p>仅返回特定字段的值，而不是全部已有的元信息字段。</p> <ul style="list-style-type: none"> 可用于降低返回的结构体大小。 不填或留空则返回所有字段。 	String Array	否

Query 节点内容：

参数名称	描述	类型	是否必选
Operation	<p>操作运算符。枚举值：</p> <ul style="list-style-type: none"> not：逻辑非。 or：逻辑或。 and：逻辑与。 lt：小于。 lte：小于等于。 gt：大于。 gte：大于等于。 eq：等于。 exist：存在性查询。 	String	是

	<ul style="list-style-type: none">• prefix：前缀查询。• match-phrase：字符串匹配查询。• nested：字段为数组时，其中同一对象内逻辑条件查询。		
SubQueries	子查询的结构体。 <ul style="list-style-type: none">• 只有当 Operations 为逻辑运算符（and、or、not或nested）时，才能设置子查询条件。• 在逻辑运算符为 and/or/not 时，其 SubQueries 内描述的所有条件需符合父级设置的 and/or/not 逻辑关系。• 在逻辑运算符为 nested 时，其父级的 Field 必须为一个数组类的字段（如：Labels），子查询条件 SubQueries 组的 Operation 必须为 and/or/not 中的一个或多个，其 Field 必须为父级 Field 的子属性。	Container Array	否
Field	字段名称。关于支持的字段，请参考 字段和操作符的支持列表 。	String	否
Value	查询的字段值。当 Operations 为逻辑运算符（and、or、not或nested）时，该字段无效。	String	否

Aggregations 节点内容：

参数名称	描述	类型	是否必选
Field	字段名称。关于支持的字段，请参考 字段和操作符的支持列表 。	String	否
Operation	聚合字段的操作符。枚举值： <ul style="list-style-type: none">• min：最小值。• max：最大值。• average：平均数。• sum：求和。• count：计数。• distinct：去重计数。• group：分组计数，按照分组计数结果从高到低排序。	String	否

响应

响应头

此接口仅返回公共响应头部，详情请参见 [公共响应头部](#) 文档。

响应体

```
{
  "Aggregations": [],
  "Files": [{
    "COSCRC64": "1111122455985959955",
    "COSStorageClass": "STANDARD",
    "CacheControl": "",
    "ContentType": "image/jpeg",
    "CreateTime": "2023-12-26T14:29:25.753167285+08:00",
    "CustomId": "002",
    "CustomLabels": {
      "age": "18",
      "level": "18"
    },
    "DatasetName": "test",
    "ETag": "\"21aeb337eb76df70dcb1e2618598****5\"",
    "FileAccessTime": "",
    "FileModifiedTime": "2023-06-07T07:20:28Z",
    "Filename": "examplebucket-1250000000/test002.jpg",
    "MediaType": "video",
    "ObjectACL": "default",
    "Size": "10000",
    "URI": "cos://examplebucket-1250000000/test002.jpg",
    "UpdateTime": "2023-12-26T14:38:11.575077683+08:00"
  }, {
    "COSCRC64": "1111125831785959934",
    "COSStorageClass": "STANDARD",
    "CacheControl": "",
    "ContentType": "image/jpeg",
    "CreateTime": "2023-12-26T11:43:18.92077679+08:00",
    "CustomId": "001",
    "CustomLabels": {
      "age": "18",
      "level": "18"
    },
    "DatasetName": "test",
    "ETag": "\"21aeb337eb76df70dcb1e2618598****\"",
    "FileModifiedTime": "2023-06-07T07:18:04Z",
    "Filename": "examplebucket-1250000000/test001.jpg",
    "MediaType": "image",
```

```
    "ObjectACL": "default",
    "Size": "20000",
    "URI": "cos://examplebucket-1250000000/test001.jpg",
    "UpdateTime": "2023-12-26T11:45:02.445509346+08:00"
  }],
  "NextToken":
"aP+BAwEBDlF1ZXJ5TmV4dFRva2VuAf+CAAEGAQpTb3J0RmllbGRzAQwAAQtTZWFyY2hBZnRl
lcgH/hAABBU9yZGVyAQwAAQVBcHBpZAEMAAEDVWluAQwAAQtEYXRhc2V0TmFtZQEMAAAAHP+
DAgEBDltdaW50ZXJmYWNlIHt9Af+EAAEQAAAj/4IBCEN1c3RvbUlkaQEGc3RyaW5nDAUAzA
wMQEEZG*****",
  "RequestId": "NWFjMzQ0MDZfOTBmYTUwXzZkZV8z*****"
}
```

响应包体具体数据内容如下：

参数名称	类型	描述
Aggregations	Container Array	聚合字段信息列表。仅在请求的 Aggregations 不为空时返回。
Files	Container Array	文件信息列表。仅在请求的 Aggregations 为空时返回。
Request Id	String	请求 ID。
NextToken	String	翻页标记。 <ul style="list-style-type: none">当文件总数大于设置的 MaxResults 时，用于翻页的 token。符合条件的文件信息未全部返回时，此参数才有值。下一次列出文件信息时将此值作为 NextToken 传入，将后续的文件信息返回。

File 节点内容：

参数名称	类型	描述
COSCRC64	String	文件 CRC64值。
COSStorage Class	String	文件存储空间类型。
CacheContro	String	指定 Object 被下载时网页的缓存行为。该字段需要设置 COS Object

l	g	HTTP 属性 Cache-Control。
ContentType	String	文件内容类型（MIME Type）。
CreateTime	String	元数据创建时间的时间戳，格式为 RFC3339Nano。
CustomId	String	该文件的自定义 ID。该文件索引到数据集后，作为该行元数据的属性存储，用于和您的业务系统进行关联、对应。您可以根据业务需求传入该值，例如将某个 URI 关联到您系统内的某个 ID。推荐传入全局唯一的值。
CustomLabels	Container	文件自定义标签列表。储存您业务自定义的键名、键值对信息，用于在查询时可以据此为筛选项进行检索。
DatasetName	String	数据集名称。
ETag	String	Object 生成时会创建相应的 ETag，ETag 用于标识一个 Object 的内容。
FileModifiedTime	String	文件最近一次修改时间的时间戳，格式为 RFC3339Nano。
Filename	String	文件路径。
MediaType	String	文件媒体类型。枚举值： <ul style="list-style-type: none"> • image：图片。 • other：其他。 • document：文档。 • archive：压缩包。 • audio：音频。 • video：视频。
ObjectACL	String	文件访问权限属性。
Size	Int	文件大小，单位为字节。
URI	String	资源标识字段，表示需要建立索引的文件地址。
UpdateTime	String	元数据修改时间的时间戳，格式为 RFC3339Nano。 创建元数据后，如果未更新过元数据，则元数据修改时间的时间戳和元数据创建时间的时间戳相同。

Objectld	String	对象唯一 ID。
ContentDisposition	String	指定 Object 被下载时的名称。需要设置 COS Object HTTP 属性 Content-Disposition。
ContentEncoding	String	指定该 Object 被下载时的内容编码格式。需要设置 COS Object HTTP 属性 Content-Encoding。
ContentLanguage	String	Object 内容使用的语言。需要设置 COS Object HTTP 属性 Content-Language。
ServerSideEncryption	String	加密算法,需要设置x-cos-server-side-encryption。
COSUserMeta	Container	COS 自定义头部。储存您业务在 cos object 上的键名、键值对信息，用于在查询时可以据此为筛选项进行检索。
COSTagging	Container	COS 自定义标签。储存您业务在 cos object 上的自定义标签的键名、键值对信息，用于在查询时可以据此为筛选项进行检索。
COSTaggingCount	Int	COS 自定义标签的数量。

Aggregations 节点内容：

参数名称	类型	描述
Field	String	聚合字段名称。
Operation	String	聚合字段的聚合操作符。
Value	Float	聚合的统计结果。
Groups	Container Array	分组聚合的结果列表。仅在请求的 Aggregations 中存在 group 类型的 Operation 时才会返回。

Groups 节点内容：

参数名称	类型	描述
Value	String	分组聚合的值。
Count	Integer	分组聚合的总个数。

实际案例

案例一：查询 size 大于1000的 jpeg 图片

请求

```
POST /datasetquery/simple HTTP/1.1
Host: <AppId>.ci.<Region>.myqcloud.com
Authorization: <Auth String>
Content-Length: <length>
Content-Type: application/json
Accept: application/json

{
  "DatasetName": "test",
  "Query": {
    "Operation": "and",
    "SubQueries": [{
      "Field": "ContentType",
      "Value": "image/jpeg",
      "Operation": "eq"
    },
    {
      "Field": "Size",
      "Value": "1000",
      "Operation": "gt"
    }
  ]
},
  "Sort": "CustomId",
  "Order": "desc",
  "MaxResults": 100
}
```

响应

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: 645
Date: Thu, 15 Jun 2023 07:59:50 GMT
Status: 200 OK
```

```
x-cos-request-id: NWFjMzQ0MDZfOTBmYTUwXzZkZV8z****

{
  "Aggregations": [],
  "Files": [{
    "COSCRC64": "1111122455985959955",
    "COSStorageClass": "STANDARD",
    "CacheControl": "",
    "ContentType": "image/jpeg",
    "CreateTime": "2023-12-26T14:29:25.753167285+08:00",
    "CustomId": "002",
    "CustomLabels": {
      "age": "18",
      "level": "18"
    },
    "DatasetName": "test",
    "ETag": "\"21aeb337eb76df70dcb1e2618598****5\"",
    "FileAccessTime": "",
    "FileModifiedTime": "2023-06-07T07:20:28Z",
    "Filename": "examplebucket-1250000000/test002.jpg",
    "MediaType": "video",
    "ObjectACL": "default",
    "Size": "10000",
    "URI": "cos://examplebucket-1250000000/test002.jpg",
    "UpdateTime": "2023-12-26T14:38:11.575077683+08:00"
  }, {
    "COSCRC64": "1111125831785959934",
    "COSStorageClass": "STANDARD",
    "CacheControl": "",
    "ContentType": "image/jpeg",
    "CreateTime": "2023-12-26T11:43:18.92077679+08:00",
    "CustomId": "001",
    "CustomLabels": {
      "age": "18",
      "level": "18"
    },
    "DatasetName": "test",
    "ETag": "\"21aeb337eb76df70dcb1e2618598****\"",
    "FileModifiedTime": "2023-06-07T07:18:04Z",
    "Filename": "examplebucket-1250000000/test001.jpg",
```

```
    "MediaType": "image",
    "ObjectACL": "default",
    "Size": "20000",
    "URI": "cos://examplebucket-1250000000/test001.jpg",
    "UpdateTime": "2023-12-26T11:45:02.445509346+08:00"
  },
  "NextToken":
"aP+BAwEBDlF1ZXJ5TmV4dFRva2VuAf+CAAEGAQpTb3J0Rml1bGRzAQwAAQtTZWFyY2hBZnRl
lcgH/hAABBU9yZGVyAQwAAQVBcHBpZAEMAAEDVWluAQwAAQtEYXRhc2V0TmFtZQEMAAAAHP+
DAgEBDltdaW50ZXJmYW50ZXJmYW50ZXJmYW50ZXJmYW50ZXJmYW50ZXJmYW50ZXJmYW50ZXJm
wMQEEZG*****",
  "RequestId": "NWFjMzQ0MDZfOTBmYTUwXzZkZV8z*****"
}
```

案例二：按文件名称聚合查询

请求

```
POST /datasetquery/simple HTTP/1.1
Host: <AppId>.ci.<Region>.myqcloud.com
Authorization: <Auth String>
Content-Length: <length>
Content-Type: application/json
Accept: application/json

{
  "Aggregations": [{
    "Field": "Filename",
    "Operation": "group"
  }],
  "DatasetName": "test",
  "MaxResults": 100
}
```

响应

```
HTTP/1.1 200 OK
Content-Type: application/json
Content-Length: 645
```

```
Date: Thu, 15 Jun 2023 07:59:50 GMT
Status: 200 OK
x-cos-request-id: NWFjMzQ0MDZfOTBmYTUwXzZkZV8z****

{
  "RequestId": "NWFjMzQ0MDZfOTBmYTUwXzZkZV8z****",
  "Aggregations": [{
    "Field": "Size",
    "Groups": [{
      "Count": 1,
      "Value": "file001"
    }, {
      "Count": 1,
      "Value": "file002"
    }],
    "Operation": "group"
  }],
  "Files": [],
  "NextToken": ""
}
```

查询结果异步导出 创建异步导出任务

最近更新时间：2025-11-13 18:09:32

功能描述

本 API 接口创建异步检索导出任务，并将检索结果，异步存放到指定的 COS 桶目录下。该接口适用于处理大规模检索请求场景，当检索结果数据庞大、无法同步返回时，用户可先发起请求，随后异步给出检索结果。

授权说明

通过子账号使用时，需要在 [授权策略](#) 的 action 中添加 `ci:CreateDatasetExport` 权限，数据万象支持的所有操作接口请参见 [action](#)。

使用限制

单账号默认并发数为1，最大排队任务数为100。

服务开通

使用该功能将默认为您开通数据万象，并绑定存储桶，无需角色授权，即可直接使用。

请求

请求示例

```
POST /datasetquery/export HTTP/1.1
Host: <AppId>.ci.<Region>.myqcloud.com
Authorization: Auth String
Content-Length: xxx
Content-Type: application/json
Accept: application/json
```

说明：

- Authorization: Auth String（详情请参见 [请求签名](#) 文档）。
- 通过子账号使用时，需要授予相关的权限，详情请参见 [授权粒度详情](#) 文档。

请求头

此接口仅使用公共请求头部，详情请参见 [公共请求头部](#) 文档。

请求体

```
{
  "Name": "fortest",
  "DatasetName": "test",
  "Query": {
    "Operation": "and",
    "SubQueries": [{
      "Field": "ContentType",
      "Value": "image/jpeg",
      "Operation": "eq"
    },
    {
      "Field": "Size",
      "Value": "1000",
      "Operation": "gt"
    }
  ]
},
  "Output": {
    "Bucket": "test-1250000000",
    "Object": "prefix001",
    "Region": "ap-beijing"
  }
}
```

请求参数

具体参数如下：

参数名称	描述	类型	是否必选
Name	新创建任务的名称，支持中文、英文、数字、一和_，长度限制128字符	String	是
DatasetName	数据集名称，同一个账户下唯一	String	是
Query	简单查询参数条件，可自嵌套	Container	否

Output	任务结果存放位置	Container	是
--------	----------	-----------	---

Container 类型 Query 的具体数据描述，参见 [简单查询](#) 接口中 Query 字段的说明。

Container 类型 Output 的具体数据描述如下：

参数名称	父节点	描述	类型	是否必选
Region	Request.Output	存储结果桶的地域	String	是
Bucket	Request.Output	存储结果的存储桶	String	是
Object	Request.Output	输出结果的文件名	String	是

响应

响应头

此接口仅返回公共响应头部，详情请参见 [公共响应头部](#) 文档。

响应体

```
{
  "Job": {
    "CreateTime": "2025-04-09T19:38:08.454310276+08:00",
    "JobId": "xxxxxxx",
    "Name": "fortest",
    "Output": {
      "Bucket": "test-1250000000",
      "Object": "prefix001",
      "Region": "ap-beijing"
    },
    "Query": {
      "Field": "",
      "Operation": "and",
      "SubQueries": [{
        "Field": "ContentType",
        "Value": "image/jpeg",
        "Operation": "eq"
      }],
      {}
    }
  }
}
```

```

        "Field": "Size",
        "Value": "1000",
        "Operation": "gt"
    },
    ],
    "Value": ""
},
"StartTime": "1970-01-01T00:00:00+08:00",
"EndTime": "1970-01-01T00:00:00+08:00",
"State": "todo"
},
"RequestId": "NjdmNjVjMjBfNzM0NjZjMT*****"
}

```

响应包体具体数据内容如下：

参数名称	类型	描述
RequestId	String	请求 ID
Job	Container	创建的任务信息

Job 节点内容：

参数名称	类型	描述
CreateTime	String	任务创建时间
JobId	String	任务 JobId，查询时需要
Name	String	任务名称
Output	Container	任务结果存放位置，和请求时传入值相同
Query	Container	简单查询参数条件，和请求时传入值相同
StartTime	String	任务启动时间
EndTime	String	任务结束时间
State	String	任务状态，包含 todo(待执行) / running(执行中) / failed(执行失败) / done(执行成功)

查询异步导出任务

最近更新时间：2025-11-13 18:09:32

功能描述

查询单个异步导出任务的状态。

授权说明

通过子账号使用时，需要在 [授权策略](#) 的 action 中添加 `ci:DescribeDatasetExport` 权限，数据万象支持的所有操作接口请参见 [action](#)。

服务开通

使用该功能将默认为您开通数据万象，并绑定存储桶，无需角色授权，即可直接使用。

使用限制

使用该接口时，请先确认相关限制条件。详情请参见 [使用限制](#)。

请求

请求示例

```
GET /datasetquery/export?datasetname=test&jobid=xxxxxxx HTTP/1.1
Host: <AppId>.ci.<Region>.myqcloud.com
Authorization: Auth String
Content-Length: xxx
Content-Type: application/json
Accept: application/json
```

❗ 说明：

- Authorization: Auth String（详情请参见 [请求签名](#) 文档）。
- 通过子账号使用时，需要授予相关的权限，详情请参见 [授权粒度详情](#) 文档。

请求头

此接口仅使用公共请求头部，详情请参见 [公共请求头部](#) 文档。

请求参数

具体参数如下：

参数名称	描述	类型	是否必选
DatasetName	数据集名称，同一个账户下唯一	String	是
JobId	创建时返回的唯一任务 Id，获取请参见 创建异步导出任务	String	是

请求体

该请求的请求体为空。

响应

响应头

此接口仅返回公共响应头部，详情请参见 [公共响应头部](#) 文档。

响应体

```
{
  "Job": {
    "CreateTime": "2025-04-09T19:38:08.454310276+08:00",
    "JobId": "xxxxxxx",
    "Name": "fortest",
    "Output": {
      "Bucket": "ci-qta-bj-1250000000",
      "Object": "prefix001",
      "Region": "ap-beijing"
    },
    "Query": {
      "Field": "",
      "Operation": "",
      "SubQueries": [],
      "Value": ""
    },
    "StartTime": "1970-01-01T00:00:00+08:00",
    "EndTime": "1970-01-01T00:00:00+08:00",
    "State": "todo"
  },
  "RequestId": "NjdmNjVjMjBfNzM0NjZjMT*****"
```

响应包体具体数据内容如下：

参数名称	类型	描述
RequestId	String	请求 ID
Job	Container	创建的任务信息

Job 节点说明：请参见 [创建异步导出任务](#) 接口响应部分的 Job 字段说明。

列出异步导出任务

最近更新时间：2025-11-13 18:09:32

功能描述

查询智能检索的异步导出任务列表。

授权说明

通过子账号使用时，需要在 [授权策略](#) 的 action 中添加 `ci:ListDatasetExport` 权限，数据万象支持的所有操作接口请参见 [action](#)。

服务开通

使用该功能将默认为您开通数据万象，并绑定存储桶，无需角色授权，即可直接使用。

使用限制

使用该接口时，请先确认相关限制条件。详情请参见 [使用限制](#)。

请求

请求示例

```
GET /datasetquery/exports?
datasetname=test&maxresults=20&nexttoken=xxxxxx HTTP/1.1
Host: <AppId>.ci.<Region>.myqcloud.com
Date: <GMT Date>
Authorization: Auth String
Accept: application/json
```

❗ 说明：

- Authorization: Auth String（详情请参见 [请求签名](#) 文档）。
- 通过子账号使用时，需要授予相关的权限，详情请参见 [授权粒度详情](#) 文档。

请求头

此接口仅使用公共请求头部，详情请参见 [公共请求头部](#) 文档。

请求参数

具体参数如下：

参数名称	描述	类型	是否必选
Dataset Name	数据集名称，同一个账户下唯一	String	是
MaxResults	返回任务列表的最大个数，取值范围为0~200。不设置此参数或者设置为0时，则默认值为100	Integer	否
NextToken	当任务总数大于设置的 MaxResults 时，用于翻页的 token。从 NextToken 开始按字典序返回任务列表。第一次调用此接口时，设置为空	String	否

请求体

该请求无请求体。

响应

响应头

此接口仅返回公共响应头部，详情请参见 [公共响应头部](#) 文档。

响应体

```
{
  "Jobs": [
    {
      "CreateTime": "2025-04-07T17:07:33+08:00",
      "EndTime": "1970-01-01T00:00:00+08:00",
      "JobId": "bd83b246138f11f0941052540006042a",
      "Name": "demo",
      "Output": {
        "Bucket": "test-1250000000",
        "Object": "test",
        "Region": "ap-beijing"
      },
      "Query": {
        "Field": "ContentType",
        "Operation": "eq",
        "SubQueries": [],
        "Value": "image/jpeg"
      }
    },
  ],
}
```

```
      "StartTime": "1970-01-01T00:00:00+08:00",
      "State": "todo"
    }
  ],
  "NextToken": "",
  "RequestId": "NjdmNjVkMjZfNzM0NjZjMTxxxxxxxxxxxx"
}
```

响应包体具体数据内容如下：

参数名称	类型	描述
RequestId	String	请求 ID
Jobs	Container Array	异步导出任务的列表
NextToken	String	当任务总数大于设置的 MaxResults 时，用于翻页的 token。下一次列出任务时以此值为 NextToken，将未返回的结果返回。当任务未全部返回时，此参数才有值

Jobs 节点说明：请参见 [创建异步导出任务](#) 接口响应部分的 Job 字段说明。

字段和操作符的支持列表

最近更新时间：2025-11-04 17:24:51

简介

通过本文您可以了解简单查询条件支持的字段（Field）和操作符（Operation），以及字段对应操作符的支持情况。

- 简单查询条件中支持使用的操作符包括 eq（等于）、gt（大于）、gte（大于等于）、lt（小于）、lte（小于等于）、nested（嵌套查询）、exist（存在性查询）、match-phrase（模糊查询）和 prefix（前缀查询）。
- 聚合操作中支持使用的操作符包括 min（最小值）、max（最大值）、average（平均数，类似于SQL中的 avg）、sum（求和）、count（计数）、distinct（去重统计行数，类似于 SQL 中的 count(distinct)，去重统计行数的计算结果是个近似值，当去重统计行数小于1万时，计算结果接近精确值，当去重统计行数达到1亿时，计算结果的误差为2%左右）、group（分组计数，当分组较大时可能会存在误差）。
- 使用自定义标签相关字段（CustomId、CustomLabels）搜索文件时，系统会按照对象标签的 Key-Value 来搜索文件。对象标签的 Key 和 Value 均不区分大小写。例如要搜索包含 Bucket = examplebucket 标签的文件，您需要设置查询条件为 {"Bucket": "examplebucket"}。

字段和操作符的支持列表

下表列出了简单查询条件和聚合操作中支持的字段，以及字段对相应操作符的支持情况。

简单查询条件

Field（字段）	类型	说明	支持的操作符	是否支持排序
COSTaggingCount	Int	COS 自定义标签的数量	eq、gt、gte、lt、lte	是
COSTagging.*	Container	COS 自定义标签。储存您业务在 cos object 上的自定义标签的键名、键值对信息，用于在查询时可以据此为筛选项进行检索	eq、gt、gte、lt、lte、prefix、exist	否
COSUserMeta.*	Container	COS 自定义头部。储存您业务在 cos object 上的键名、键值对信息，用于在查询时可以据此为筛选项进行检索	eq、gt、gte、lt、lte、prefix、exist	否
CreateTime	String	元数据创建时间的时间戳，格式为 RFC3339Nano	eq、gt、gte、lt、lte	是

CustomId	String	该文件的自定义 ID。该文件索引到数据集后，作为该行元数据的属性存储，用于和您的业务系统进行关联、对应。您可以根据业务需求传入该值，例如将某个 URI 关联到您系统内的某个 ID。推荐传入全局唯一的值	eq、gt、gte、lt、lte、prefix	是
CustomLabels.*	Container	文件自定义标签列表。储存您业务自定义的键名、键值对信息，用于在查询时可以据此为筛选项进行检索	eq、gt、gte、lt、lte、prefix、exist	否
ETag	String	Object 生成时会创建相应的 ETag，ETag 用于标识一个 Object 的内容	eq、gt、gte、lt、lte、prefix	是
FileModifiedTime	String	文件最近一次修改时间的时间戳，格式为 RFC3339Nano	eq、gt、gte、lt、lte	是
Filename	String	文件路径	eq、gt、gte、lt、lte、match-phrase、prefix	是
MediaType	String	文件媒体类型 枚举值： <ul style="list-style-type: none"> • image: 图片 • other: 其他 • document: 文档 • archive: 压缩包 • audio: 音频 • video: 视频 	eq、gt、gte、lt、lte、prefix	是
ObjectACL	String	文件访问权限属性	eq、gt、gte、lt、lte	是
ObjectId	String	对象唯一 ID	eq、gt、gte、lt、lte、prefix	是
ServerSideEncryption	String	加密算法，需要设置 x-cos-server-side-encryption	eq、gt、gte、lt、lte、prefix	是
Size	Int	文件大小，单位为字节	eq、gt、gte、lt、lte	是

UpdateTime	String	元数据修改时间的时间戳，格式为 RFC3339Nano 创建元数据后，如果未更新过元数据，则元数据修改时间的时间戳和元数据创建时间的时间戳相同	eq、gt、gte、lt、lte	是
URI	String	资源标识字段，表示需要建立索引的文件地址	eq、gt、gte、lt、lte、prefix	是

聚合操作

Field（字段）	类型	说明	支持的操作符	是否支持排序
ContentDisposition	String	指定 Object 被下载时的名称。需要设置 COS Object HTTP 属性 Content-Disposition	count、distinct、group	是
ContentEncoding	String	指定该 Object 被下载时的内容编码格式。需要设置 COS Object HTTP 属性 Content-Encoding	count、distinct、group	是
ContentLanguage	String	Object 内容使用的语言。需要设置 COS Object HTTP 属性 Content-Language	count、distinct、group	是
ContentType	String	文件内容类型（MIME Type）	count、distinct、group	是
COSCRC64	String	文件 CRC64 值	count、distinct、group	是
COSStorageClass	String	文件存储空间类型	count、distinct、group	是
COSTagging.*	Container	COS 自定义标签。储存您业务在 cos object 上的自定义标签的键名、键值对信息，用于在查询时可以据此为筛选项进行检索	-	否
COSTaggingCount	Int	COS 自定义标签的数量	min、max、average、sum、count、distinct、group	是

COSUserMeta.*	Container	COS 自定义头部。储存您业务在 cos object 上的键名、键值对信息，用于在查询时可以据此为筛选项进行检索	—	否
CreateTime	String	元数据创建时间的时间戳，格式为 RFC3339Nano	count、distinct、group	是
CustomId	String	该文件的自定义 ID。该文件索引到数据集后，作为该行元数据的属性存储，用于和您的业务系统进行关联、对应。您可以根据业务需求传入该值，例如将某个 URI 关联到您系统内的某个 ID。推荐传入全局唯一的值	count、distinct、group	是
CustomLabels.*	Container	文件自定义标签列表。储存您业务自定义的键名、键值对信息，用于在查询时可以据此为筛选项进行检索	—	否
ETag	String	Object 生成时会创建相应的 ETag，ETag 用于标识一个 Object 的内容	count、distinct、group	是
FileModifiedTime	String	文件最近一次修改时间的时间戳，格式为 RFC3339Nano	count、distinct、group	是
Filename	String	文件路径	count、distinct、group	是
MediaType	String	文件媒体类型 枚举值： <ul style="list-style-type: none"> • image: 图片 • other: 其他 • document: 文档 • archive: 压缩包 • audio: 音频 • video: 视频 	count、distinct、group	是
ObjectACL	String	文件访问权限属性	count、distinct、group	是
ObjectId	String	对象唯一 ID	count、distinct、group	是

ServerSideEncryption	String	加密算法，需要设置 x-cos-server-side-encryption	count、distinct、group	是
Size	Int	文件大小，单位为字节	min、max、average、sum、count、distinct、group	是
UpdateTime	String	元数据修改时间的时间戳，格式为 RFC3339Nano 创建元数据后，如果未更新过元数据，则元数据修改时间的时间戳和元数据创建时间的时间戳相同	count、distinct、group	是
URI	String	资源标识字段，表示需要建立索引的文件地址	count、distinct、group	是

向量检索

向量检索概述

最近更新时间：2025-08-29 15:03:52

向量检索从 COS 对象中提取特征信息，使用户能够基于语义对 COS 对象进行检索，为非结构化数据的分析与检索提供了强大的支持。

使用场景

网盘相册

向量检索可将每张图片转化为向量表示，可满足用户拍照搜图，自然语言搜索需求，实现更加智能和精准的图片搜索体验，提升用户检索效率和满意度。

媒资管理

可以将媒资内容如文本、图像、视频等非结构化数据表示为向量。用户检索相关内容时，可进行以文搜图、以图搜图、以视频搜视频、人脸搜索等跨模态检索操作，提高了媒资内容的管理和分析能力。

操作步骤

1. 登录 [对象存储控制台](#)。
2. 在左侧导航栏中，选择**智能检索 > 数据检索**。
3. 选择检索场景**图像检索**或**人脸检索**。
4. 选择需要进行检索的数据集。
5. 输入与检索内容相关的文字信息或选择需要检索的图片。
6. 展开高级设置，可调整返回结果数量与相似度。
7. 单击**查询**，获取检索结果。

对象存储

概览

存储桶列表

收藏路径

运维中心

使用统计

资源包管理

存储+

智能检索

数据检索

数据集管理

模版管理

用量统计

智能工具箱

数据工作流

批量处理

云原生数据湖存储

数据检索使用指引

1 选择检索场景

2 选择数据集

3 输入检索条件

4 查看检索结果

基础信息检索 图像检索 人脸检索

选择数据集


te

高级设置

番茄

查询

查询结果



相似度：76

复制URI

图像检索

最近更新时间：2025-11-17 15:20:22

功能描述

支持输入自然语言文本或者图片，在指定数据集中检索出符合输入的文本语义或图片内容相似的图片。

授权说明

通过子账号使用时，需要在 [授权策略](#) 的 action 中添加 `ci:DatasetImageSearch` 权限。数据万象支持的所有操作接口请参见 [CI action](#)。

使用限制

返回相关图片数量最大为100。

费用说明

有关图像检索的费用，请参见 [智能检索费用](#)。

请求

请求示例

```
POST /datasetquery/imagesearch HTTP/1.1
Host: <AppId>.ci.<Region>.myqcloud.com
Authorization: Auth String
Content-Length: xxx
Content-Type: application/json
Accept: application/json
```



说明：

Authorization: Auth String，详情请参见 [请求签名](#) 文档。

请求头

此接口仅使用公共请求头部，详情请参见 [公共请求头部](#) 文档。

请求体

```
{
```

```
"DatasetName": "test",
"Mode": "pic",
"URI": "cos://facesearch-125872xxxx/huge_base.jpg",
"Limit": 10,
"MatchThreshold": 1
}
```

请求参数

参数名称	描述	类型	是否必选
DatasetName	数据集名称，同一个账户下唯一。该场景下仅支持选择绑定图像检索模板的数据集。可通过 控制台 或 接口 查询。	String	是
Mode	指定检索的输入类型，有效值为： <ul style="list-style-type: none">pic：表示输入图片进行以图搜图的检索。text：表示输入文本进行检索，支持输入自然语言，例如“包含一棵大树的图片”。 默认值为 pic。	String	否
URI	资源标识字段，表示需要建立索引的文件地址(Mode 为 pic 时必选)。	String	否
Limit	返回相关图片的数量，默认值为10，最大值为100。	Integer	否
Text	检索语句，检索方式为 text 时必填，最多支持60个字符(Mode 为 text 时必选)。 示例：包含一棵大树的图片	String	否
MatchThreshold	限制返回图片的最低相关度分数，只有超过 MatchThreshold 值的图片才会返回。默认值为0，推荐值为80。 例如：设置 MatchThreshold 的值为80，则检索结果中仅会返回相关度分数大于等于80分的图片。	Integer	否

响应

响应头

此接口仅返回公共响应头部，详情请参见 [公共响应头部](#) 文档。

响应体

```
{
```

```
"ImageResult": [{
  "URI": "cos://facesearch-125872xxxx/huge_base.jpg",
  "Score": 99
}],
"RequestId": "NjYwYzEwYjhfNGxxxxxxxJfMjcwNl82"
```

响应包体具体数据内容如下：

参数名称	类型	描述
ImageResult	Container Array	图像检索识别结果信息列表。
RequestId	String	请求 ID。

ImageResult 节点内容：

参数名称	类型	描述
URI	String	资源标识字段，表示需要建立索引的文件地址。
Score	Integer	图片相关度分数，分数越高表示与检索条件越相符。

人脸检索

最近更新时间：2025-08-29 15:03:52

功能描述

支持用一张待检测的人脸图片，从数据集中检索出与最相似的前 N 张人脸图片。

授权说明

通过子账号使用时，需要在 [授权策略](#) 的 action 中添加 `ci:DatasetFaceSearch` 权限。数据万象支持的所有操作接口请参见 [CI action](#)。

使用限制

- 每次检索，可识别图片中的最多 10 张人脸。
- 每张人脸最多可返回相关人脸数量为10，即一张人脸图片最多可返回相关人脸图片数量为100。

费用说明

有关人脸检索的费用，请参见 [智能检索费用](#)。

请求

请求示例

```
POST /datasetquery/facesearch HTTP/1.1
Host: <AppId>.ci.<Region>.myqcloud.com
Authorization: Auth String
Content-Length: xxx
Content-Type: application/json
Accept: application/json
```



说明：

Authorization: Auth String，详情请参见 [请求签名](#) 文档。

请求头

此接口仅使用公共请求头部，详情请参见 [公共请求头部](#) 文档。

请求体

```
{
  "DatasetName": "test",
  "URI": "cos://examplebucket-1250000000/test.jpg",
  "MaxFaceNum": 1,
  "Limit": 10,
  "MatchThreshold": 10
}
```

请求参数

参数名称	描述	类型	是否必选
DatasetName	数据集名称，同一个账户下唯一。该场景下仅支持选择绑定人脸检索模板的数据集。可通过 控制台 或 接口 查询。	String	是
URI	资源标识字段，表示需要建立索引的文件地址。	String	是
MaxFaceNum	输入图片中检索的人脸数量，默认值为1(传0或不传采用默认值)，最大值为10。	Integer	否
Limit	检索的每张人脸返回相关人脸数量，默认值为10，最大值为100。	Integer	否
MatchThreshold	限制返回人脸的最低相关度分数，只有超过 MatchThreshold 值的人脸才会返回。默认值为0，推荐值为80。 例如：设置 MatchThreshold 的值为80，则检索结果中仅会返回相关度分数大于等于80分的人脸。	Integer	否

响应

响应头

此接口仅返回公共响应头部，详情请参见 [公共响应头部](#) 文档。

响应体

```
{
  "FaceResult": [{
    "FaceInfos": [{
      "PersonId": "xxxxxx",
      "FaceBoundary": {
        "Height": 264,
```

```
        "Width": 203,
        "Left": 353,
        "Top": 90
    },
    "FaceId": "80c10056-1d40-418b-9f4f-dabf8e8cc349",
    "Score": 76,
    "URI": "cos://facesearch-1258726280/huge_hezao.webp"
  }],
  "InputFaceBoundary": {
    "Height": 545,
    "Width": 401,
    "Left": 737,
    "Top": 191
  }
}],
  "RequestId": "NjYxNTMyY2JfNGQ2ODk0MGJfNzAzZl81"
}
```

响应包体具体数据内容如下：

参数名称	类型	描述
FaceResult	Container Array	人脸检索识别结果信息列表。
RequestId	String	请求 ID。

FaceResult 节点内容：

参数名称	类型	描述
FaceInfos	Container Array	相关人脸信息列表。
InputFaceBoundary	Container	输入图片的人脸框位置。

FaceInfos 节点内容：

参数名称	类型	描述
PersonId	String	自定义人物ID。
FaceBoundary	Container	相关人脸框位置。

FaceId	String	人脸ID。
Score	Integer	相关人脸匹配得分。
URI	String	资源标识字段，表示需要建立索引的文件地址。

FaceBoundary 节点内容：

参数名称	类型	描述
Height	Integer	人脸高度。
Width	Integer	人脸宽度。
Left	Integer	人脸框左上角横坐标。
Top	Integer	人脸框左上角纵坐标。

InputFaceBoundary 节点内容：

参数名称	类型	描述
Height	Integer	人脸高度。
Width	Integer	人脸宽度。
Left	Integer	人脸框左上角横坐标。
Top	Integer	人脸框左上角纵坐标。