

# CECSTACK 5.3.0

## AI Space OpenAPI 参考

---

文档密级：公开

文档版本：01


发布日期：2025-02-13

### 【版权声明】

版权所有 © 中电云计算技术有限公司 2025。保留一切权利。

本文档的版权归中电云计算技术有限公司所有。非经中电云计算技术有限公司书面许可，任何人不得以包括通过程序或设备监视、复制、传播、展示、镜像、上载、下载、摘编等方式或以其他方式擅自使用本文档的任何内容。

### 【商标声明】

 中国电子云 和本文档所示其他中电云计算技术有限公司及/或其他关联公司的商标均为中电云计算技术有限公司及/或其关联公司所有。未经中电云计算技术有限公司及/或其关联公司书面许可，任何人不得以任何形式使用，也不得向他人表明您有权展示、使用或做其他处理。如您有宣传、展示等任何使用需要，您必须取得中电云计算技术有限公司及/或其关联公司事先书面授权。

本文档中出现的其他公司的商标或注册商标，由各自的所有人拥有。

### 【注意】

您购买的产品、服务或特性等应以中电云计算技术有限公司商业合同中的约定为准，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，中电云计算技术有限公司对本文档内容不做任何明示或默示的声明或保证。

由于产品版本升级或其他原因，本文档内容可能会不定期进行更新。本文档仅作为使用指导，其中的陈述、信息或建议等均不构成任何明示或暗示的担保。

# 前言

## 概述

本文档主要介绍如何使用 AISpace 的 OpenAPI 接口。

## 读者对象





本文档适用于以下读者：

- 维护工程师
- 技术支持工程师
- 系统管理员

## 本书约定

### 符号标志约定

本书采用各种醒目标志来表示在操作过程中应该特别注意的地方，这些标志的意义如下：

 警告	该标志后的注释需给予格外关注，不当的操作可能会对人身造成伤害。
 注意	提醒操作中应注意的事项，不当的操作可能会导致数据丢失或者设备损坏。 “注意”不涉及人身伤害。
 说明	对正文的重点信息进行必要的补充说明。 “说明”不是安全警示信息，不涉及人身、设备及环境伤害信息。
 提示	配置、操作、或使用产品的技巧、窍门。

## 修订记录

文档版本	发布时间	修订说明
01	2025-02-13	第一次正式发布。

# 目 录

<b>1 公共的 Header 参数</b> .....	<b>1</b>
1.1 简要描述 .....	1
1.2 请求参数 .....	1
<b>2 开发机 API 参考</b> .....	<b>2</b>
2.1 产品功能简介 .....	2
2.2 Notebook 创建接口.....	2
2.2.1 简要描述 .....	2
2.2.2 接口约束 .....	2
2.2.3 URL 格式.....	2
2.2.4 请求方式 .....	2
2.2.5 请求参数 .....	2
2.2.6 返回参数 .....	4
2.2.7 示例 .....	4
2.3 Notebook 编辑接口.....	5
2.3.1 简要描述 .....	5
2.3.2 接口约束 .....	5
2.3.3 URL 格式.....	5
2.3.4 请求方式 .....	6
2.3.5 请求参数 .....	6
2.3.6 返回参数 .....	7
2.3.7 示例 .....	7
2.4 停止 Notebook 接口.....	8
2.4.1 简要描述 .....	8
2.4.2 接口约束 .....	8
2.4.3 URL 格式.....	8
2.4.4 请求方式 .....	8
2.4.5 请求参数 .....	8
2.4.6 返回参数 .....	9
2.4.7 示例 .....	9
2.5 重置 Notebook 接口.....	10
2.5.1 简要描述 .....	10
2.5.2 接口约束 .....	10
2.5.3 URL 格式.....	10

2.5.4 请求方式 .....	10
2.5.5 请求参数 .....	10
2.5.6 返回参数 .....	11
2.5.7 示例 .....	11
2.6 Notebook 删除接口 .....	11
2.6.1 简要描述 .....	11
2.6.2 接口约束 .....	11
2.6.3 URL 格式 .....	11
2.6.4 请求方式 .....	12
2.6.5 请求参数 .....	12
2.6.6 返回参数 .....	12
2.6.7 示例 .....	13
<b>3 在线服务 API 参考 .....</b>	<b>14</b>
3.1 产品功能简介 .....	14
3.2 在线服务创建接口 .....	14
3.2.1 简要描述 .....	14
3.2.2 接口约束 .....	14
3.2.3 URL 格式 .....	14
3.2.4 请求方式 .....	14
3.2.5 请求参数 .....	15
3.2.6 返回参数 .....	16
3.2.7 示例 .....	16
3.3 在线服务编辑接口 .....	17
3.3.1 简要描述 .....	17
3.3.2 接口约束 .....	17
3.3.3 URL 格式 .....	17
3.3.4 请求方式 .....	17
3.3.5 请求参数 .....	17
3.3.6 返回参数 .....	19
3.3.7 示例 .....	19
3.4 在线服务发布接口 .....	20
3.4.1 简要描述 .....	20
3.4.2 接口约束 .....	20
3.4.3 URL 格式 .....	20
3.4.4 请求方式 .....	20
3.4.5 请求参数 .....	20

3.4.6 返回参数 .....	21
3.4.7 示例 .....	21
3.5 在线服务下线接口 .....	21
3.5.1 简要描述 .....	21
3.5.2 接口约束 .....	21
3.5.3 URL 格式 .....	22
3.5.4 请求方式 .....	22
3.5.5 请求参数 .....	22
3.5.6 返回参数 .....	22
3.5.7 示例 .....	22
3.6 在线服务调试接口 .....	23
3.6.1 简要描述 .....	23
3.6.2 接口约束 .....	23
3.6.3 URL 格式 .....	23
3.6.4 请求方式 .....	24
3.6.5 请求参数 .....	24
3.6.6 返回参数 .....	24
3.6.7 示例 .....	24
3.7 在线服务清理接口 .....	25
3.7.1 简要描述 .....	25
3.7.2 接口约束 .....	25
3.7.3 URL 格式 .....	25
3.7.4 请求方式 .....	25
3.7.5 请求参数 .....	25
3.7.6 返回参数 .....	26
3.7.7 示例 .....	26
<b>4 公共资源池 API 参考 .....</b>	<b>27</b>
4.1 产品功能简介 .....	27
4.2 购买资源包接口 .....	27
4.2.1 简要描述 .....	27
4.2.2 接口约束 .....	27
4.2.3 URL 格式 .....	27
4.2.4 请求方式 .....	27
4.2.5 请求参数 .....	27
4.2.6 返回参数 .....	29
4.2.7 示例 .....	29

4.3 资源包列表接口 .....	30
4.3.1 简要描述 .....	30
4.3.2 接口约束 .....	30
4.3.3 URL 格式 .....	30
4.3.4 请求方式 .....	30
4.3.5 请求参数 .....	30
4.3.6 返回参数 .....	31
4.3.7 示例 .....	33
4.4 资源包详情接口 .....	34
4.4.1 简要描述 .....	34
4.4.2 接口约束 .....	34
4.4.3 URL 格式 .....	34
4.4.4 请求方式 .....	34
4.4.5 请求参数 .....	34
4.4.6 返回参数 .....	35
4.4.7 示例 .....	36
<b>5 存储资源 API 参考 .....</b>	<b>39</b>
5.1 产品功能简介 .....	39
5.2 存储资源购买接口 .....	39
5.2.1 简要描述 .....	39
5.2.2 接口约束 .....	39
5.2.3 URL 格式 .....	39
5.2.4 请求方式 .....	39
5.2.5 请求参数 .....	39
5.2.6 返回参数 .....	41
5.2.7 示例 .....	41
5.3 存储资源列表接口 .....	42
5.3.1 简要描述 .....	42
5.3.2 接口约束 .....	42
5.3.3 URL 格式 .....	42
5.3.4 请求方式 .....	42
5.3.5 请求参数 .....	42
5.3.6 返回参数 .....	43
5.3.7 示例 .....	45
<b>6 专属资源池 API 参考 .....</b>	<b>47</b>
6.1 产品功能简介 .....	47

<b>6.2 专属资源池购买接口</b> .....	47
<b>6.2.1 简要描述</b> .....	47
<b>6.2.2 接口约束</b> .....	47
<b>6.2.3 URL 格式</b> .....	47
<b>6.2.4 请求方式</b> .....	47
<b>6.2.5 请求参数</b> .....	47
<b>6.2.6 返回参数</b> .....	50
<b>6.2.7 示例</b> .....	50
<b>6.3 专属资源池退订接口</b> .....	51
<b>6.3.1 简要描述</b> .....	51
<b>6.3.2 接口约束</b> .....	51
<b>6.3.3 URL 格式</b> .....	51
<b>6.3.4 请求方式</b> .....	51
<b>6.3.5 请求参数</b> .....	51
<b>6.3.6 返回参数</b> .....	52
<b>6.3.7 示例</b> .....	53
<b>7 大模型 API 参考</b> .....	54
<b>7.1 产品功能简介</b> .....	54
<b>7.2 大模型预训练、调优(SFT、DPO、RLHF)创建接口</b> .....	54
<b>7.2.1 简要描述</b> .....	54
<b>7.2.2 接口约束</b> .....	54
<b>7.2.3 URL 格式</b> .....	54
<b>7.2.4 请求方式</b> .....	55
<b>7.2.5 请求参数</b> .....	55
<b>7.2.6 返回参数</b> .....	56
<b>7.2.7 示例</b> .....	57
<b>7.3 大模型预训练、调优(SFT、DPO、RLHF)编辑接口</b> .....	60
<b>7.3.1 简要描述</b> .....	60
<b>7.3.2 接口约束</b> .....	60
<b>7.3.3 URL 格式</b> .....	60
<b>7.3.4 请求方式</b> .....	60
<b>7.3.5 请求参数</b> .....	60
<b>7.3.6 返回参数</b> .....	62
<b>7.3.7 示例</b> .....	63
<b>7.4 大模型预训练、调优(SFT、DPO、RLHF)详情接口</b> .....	65
<b>7.4.1 简要描述</b> .....	65



7.4.2 接口约束 .....	66
7.4.3 URL 格式 .....	66
7.4.4 请求方式 .....	66
7.4.5 请求参数 .....	66
7.4.6 返回参数 .....	66
7.4.7 示例 .....	69
7.5 大模型预训练、调优(SFT、DPO、RLHF)列表接口 .....	73
7.5.1 简要描述 .....	73
7.5.2 接口约束 .....	73
7.5.3 URL 格式 .....	73
7.5.4 请求方式 .....	73
7.5.5 请求参数 .....	73
7.5.6 返回参数 .....	74
7.5.7 示例 .....	75
7.6 大模型预训练、调优(SFT、DPO、RLHF)运行接口 .....	77
7.6.1 简要描述 .....	77
7.6.2 接口约束 .....	77
7.6.3 URL 格式 .....	77
7.6.4 请求方式 .....	78
7.6.5 请求参数 .....	78
7.6.6 返回参数 .....	78
7.6.7 示例 .....	78
7.7 大模型预训练、调优(SFT、DPO、RLHF)停止接口 .....	78
7.7.1 简要描述 .....	78
7.7.2 接口约束 .....	79
7.7.3 URL 格式 .....	79
7.7.4 请求方式 .....	79
7.7.5 请求参数 .....	79
7.7.6 返回参数 .....	80
7.7.7 示例 .....	80
7.8 大模型预训练、调优(SFT、DPO、RLHF)删除接口 .....	80
7.8.1 简要描述 .....	80
7.8.2 接口约束 .....	80
7.8.3 URL 格式 .....	80
7.8.4 请求方式 .....	81
7.8.5 请求参数 .....	81

7.8.6 返回参数 .....	81
7.8.7 示例 .....	82
7.9 大模型评估任务创建接口 .....	85
7.9.1 简要描述 .....	85
7.9.2 接口约束 .....	85
7.9.3 URL 格式 .....	85
7.9.4 请求方式 .....	85
7.9.5 请求参数 .....	85
7.9.6 返回参数 .....	87
7.9.7 示例 .....	87
7.10 大模型评估任务编辑接口 .....	88
7.10.1 简要描述 .....	88
7.10.2 接口约束 .....	88
7.10.3 URL 格式 .....	88
7.10.4 请求方式 .....	89
7.10.5 请求参数 .....	89
7.10.6 返回参数 .....	90
7.10.7 示例 .....	91
7.11 大模型评估任务运行接口 .....	92
7.11.1 简要描述 .....	92
7.11.2 接口约束 .....	92
7.11.3 URL 格式 .....	92
7.11.4 请求方式 .....	92
7.11.5 请求参数 .....	92
7.11.6 返回参数 .....	93
7.11.7 示例 .....	93
7.12 大模型评估任务停止接口 .....	93
7.12.1 简要描述 .....	93
7.12.2 接口约束 .....	93
7.12.3 URL 格式 .....	93
7.12.4 请求方式 .....	94
7.12.5 请求参数 .....	94
7.12.6 返回参数 .....	94
7.12.7 示例 .....	94
7.13 大模型评估任务删除接口 .....	94
7.13.1 简要描述 .....	94

7.13.2	接口约束 .....	95
7.13.3	URL 格式 .....	95
7.13.4	请求方式 .....	95
7.13.5	请求参数 .....	95
7.13.6	返回参数 .....	95
7.13.7	示例 .....	96
7.14	大模型评估任务详情接口 .....	96
7.14.1	简要描述 .....	96
7.14.2	接口约束 .....	96
7.14.3	URL 格式 .....	97
7.14.4	请求方式 .....	97
7.14.5	请求参数 .....	97
7.14.6	返回参数 .....	97
7.14.7	示例 .....	98
7.15	大模型任务创建接口 .....	100
7.15.1	简要描述 .....	100
7.15.2	接口约束 .....	100
7.15.3	URL 格式 .....	100
7.15.4	请求方式 .....	100
7.15.5	请求参数 .....	100
7.15.6	返回参数 .....	101
7.15.7	示例 .....	101
7.16	大模型任务删除接口 .....	102
7.16.1	简要描述 .....	102
7.16.2	接口约束 .....	102
7.16.3	URL 格式 .....	102
7.16.4	请求方式 .....	102
7.16.5	请求参数 .....	102
7.16.6	返回参数 .....	103
7.16.7	示例 .....	104
<b>8</b>	<b>镜像 API 参考 .....</b>	<b>106</b>
8.1	产品功能简介 .....	106
8.2	镜像构建接口 .....	106
8.2.1	简要描述 .....	106
8.2.2	接口约束 .....	106
8.2.3	URL 格式 .....	106

8.2.4	请求方式 .....	106
8.2.5	请求参数 .....	106
8.2.6	返回参数 .....	107
8.2.7	示例 .....	108
8.3	镜像构建删除接口 .....	109
8.3.1	简要描述 .....	109
8.3.2	接口约束 .....	109
8.3.3	URL 格式 .....	109
8.3.4	请求方式 .....	109
8.3.5	请求参数 .....	110
8.3.6	返回参数 .....	110
8.3.7	示例 .....	111
8.4	镜像列表查询接口 .....	112
8.4.1	简要描述 .....	112
8.4.2	接口约束 .....	112
8.4.3	URL 格式 .....	112
8.4.4	请求方式 .....	112
8.4.5	请求参数 .....	112
8.4.6	返回参数 .....	113
8.4.7	示例 .....	113
8.5	自定义镜像上传接口 .....	119
8.5.1	简要描述 .....	119
8.5.2	接口约束 .....	119
8.5.3	URL 格式 .....	120
8.5.4	请求方式 .....	120
8.5.5	请求参数 .....	120
8.5.6	返回参数 .....	120
8.5.7	示例 .....	121
8.6	镜像推送接口 .....	121
8.6.1	简要描述 .....	121
8.6.2	接口约束 .....	121
8.6.3	URL 格式 .....	121
8.6.4	请求方式 .....	122
8.6.5	请求参数 .....	122
8.6.6	返回参数 .....	122
8.6.7	示例 .....	122

<b>9 项目空间参考</b> .....	<b>124</b>
<b>9.1 项目空间创建接口</b> .....	<b>124</b>
<b>9.1.1 简要描述</b> .....	<b>124</b>
<b>9.1.2 接口约束</b> .....	<b>124</b>
<b>9.1.3 URL 格式</b> .....	<b>124</b>
<b>9.1.4 请求方式</b> .....	<b>124</b>
<b>9.1.5 请求参数</b> .....	<b>124</b>
<b>9.1.6 返回参数</b> .....	<b>125</b>
<b>9.1.7 示例</b> .....	<b>125</b>
<b>9.2 关联资源接口</b> .....	<b>125</b>
<b>9.2.1 简要描述</b> .....	<b>125</b>
<b>9.2.2 接口约束</b> .....	<b>125</b>
<b>9.2.3 URL 格式</b> .....	<b>125</b>
<b>9.2.4 请求方式</b> .....	<b>126</b>
<b>9.2.5 请求参数</b> .....	<b>126</b>
<b>9.2.6 返回参数</b> .....	<b>126</b>
<b>9.2.7 示例</b> .....	<b>126</b>
<b>9.3 删除项目接口</b> .....	<b>127</b>
<b>9.3.1 简要描述</b> .....	<b>127</b>
<b>9.3.2 接口约束</b> .....	<b>127</b>
<b>9.3.3 URL 格式</b> .....	<b>127</b>
<b>9.3.4 请求方式</b> .....	<b>127</b>
<b>9.3.5 请求参数</b> .....	<b>127</b>
<b>9.3.6 返回参数</b> .....	<b>127</b>
<b>9.3.7 示例</b> .....	<b>128</b>
<b>9.4 移除资源</b> .....	<b>128</b>
<b>9.4.1 简要描述</b> .....	<b>128</b>
<b>9.4.2 接口约束</b> .....	<b>128</b>
<b>9.4.3 URL 格式</b> .....	<b>128</b>
<b>9.4.4 请求方式</b> .....	<b>129</b>
<b>9.4.5 请求参数</b> .....	<b>129</b>
<b>9.4.6 返回参数</b> .....	<b>129</b>
<b>9.4.7 示例</b> .....	<b>129</b>
<b>9.5 查询项目</b> .....	<b>130</b>
<b>9.5.1 简要描述</b> .....	<b>130</b>
<b>9.5.2 接口约束</b> .....	<b>130</b>

9.5.3 URL 格式 .....	130
9.5.4 请求方式 .....	130
9.5.5 请求参数 .....	130
9.5.6 返回参数 .....	131
9.5.7 示例 .....	131
<b>10 数据集管理 .....</b>	<b>133</b>
<b>10.1 数据集创建 .....</b>	<b>133</b>
10.1.1 接口约束 .....	133
10.1.2 URL 格式 .....	133
10.1.3 请求方式 .....	133
10.1.4 请求参数 .....	133
10.1.5 返回参数 .....	134
10.1.6 示例 .....	134
<b>10.2 上传文件 .....</b>	<b>135</b>
10.2.1 简要描述 .....	135
10.2.2 接口约束 .....	135
10.2.3 URL 格式 .....	135
10.2.4 请求方式 .....	135
10.2.5 请求参数 .....	135
10.2.6 返回参数 .....	135
10.2.7 示例 .....	136
<b>10.3 查询数据集 .....</b>	<b>136</b>
10.3.1 简要描述 .....	136
10.3.2 接口约束 .....	136
10.3.3 URL 格式 .....	136
10.3.4 请求方式 .....	136
10.3.5 请求参数 .....	137
10.3.6 返回参数 .....	137
10.3.7 示例 .....	138
<b>10.4 删除数据集 .....</b>	<b>139</b>
10.4.1 简要描述 .....	139
10.4.2 接口约束 .....	139
10.4.3 URL 格式 .....	139
10.4.4 请求方式 .....	139
10.4.5 请求参数 .....	139
10.4.6 返回参数 .....	139

10.4.7 示例 .....	140
<b>11 训练作业 API 参考</b> .....	<b>141</b>
11.1 产品功能简介 .....	141
11.2 训练作业创建接口 .....	141
11.2.1 简要描述 .....	141
11.2.2 接口约束 .....	141
11.2.3 URL 格式 .....	141
11.2.4 请求方式 .....	141
11.2.5 请求参数 .....	141
11.2.6 返回参数 .....	142
11.2.7 示例 .....	143
11.3 训练作业编辑接口 .....	144
11.3.1 简要描述 .....	144
11.3.2 接口约束 .....	144
11.3.3 URL 格式 .....	144
11.3.4 请求方式 .....	145
11.3.5 请求参数 .....	145
11.3.6 返回参数 .....	146
11.3.7 示例 .....	146
11.4 训练作业运行接口 .....	148
11.4.1 简要描述 .....	148
11.4.2 接口约束 .....	148
11.4.3 URL 格式 .....	149
11.4.4 请求方式 .....	149
11.4.5 请求参数 .....	149
11.4.6 返回参数 .....	149
11.4.7 示例 .....	149
11.5 训练作业停止接口 .....	150
11.5.1 简要描述 .....	150
11.5.2 接口约束 .....	150
11.5.3 URL 格式 .....	150
11.5.4 请求方式 .....	150
11.5.5 请求参数 .....	150
11.5.6 返回参数 .....	151
11.5.7 示例 .....	151
11.6 训练作业删除接口 .....	151

11.6.1 简要描述 .....	151
11.6.2 接口约束 .....	151
11.6.3 URL 格式 .....	151
11.6.4 请求方式 .....	152
11.6.5 请求参数 .....	152
11.6.6 返回参数 .....	152
11.6.7 10.6.7 示例 .....	152
<b>12 实验管理 API 参考 .....</b>	<b>153</b>
12.1 产品功能简介 .....	153
12.2 实验管理创建接口 .....	153
12.2.1 简要描述 .....	153
12.2.2 接口约束 .....	153
12.2.3 URL 格式 .....	153
12.2.4 请求方式 .....	153
12.2.5 请求参数 .....	153
12.2.6 返回参数 .....	154
12.2.7 示例 .....	154
12.3 实验管理删除接口 .....	155
12.3.1 简要描述 .....	155
12.3.2 接口约束 .....	155
12.3.3 URL 格式 .....	155
12.3.4 请求方式 .....	155
12.3.5 请求参数 .....	155
12.3.6 返回参数 .....	156
12.3.7 示例 .....	156
<b>13 接口集成参考 .....</b>	<b>157</b>
13.1 端到端功能简介 .....	157
13.2 Notebook 创建 .....	157
13.2.1 简要描述 .....	157
13.2.2 接口约束 .....	157
13.2.3 端到端业务流程 .....	157
13.3 大模型调优 SFT 训练任务创建 .....	158
13.3.1 简要描述 .....	158
13.3.2 接口约束 .....	158
13.3.3 端到端业务流程 .....	158



# 1 公共的 Header 参数

## 1.1 简要描述

此部分对于各接口用到的 Header 参数进行统一说明，涉及签名计算、认证鉴权等相关的参数详见《公共部分 OpenAPI 参考》。

## 1.2 请求参数

参数	是否必选	参数类型	示例	描述
x-regionid	是	String	cn-beijing-1	区域ID，请通过查询区域信息的接口进行获取

# 2 开发机 API 参考

## 2.1 产品功能简介

开发机是 AI 开发平台为机器学习开发者提供的在线编译、调试代码和模型开发的模块。开发机支持使用 WebIDE 在线开发，或通过 SSH 连接开发机远程开发，也提供了持久化的对象存储服务和共享文件系统用于存储开发中的数据。通过开发机环境，开发人员可以进行交互式建模，直接在开发机环境下编写数据处理和模型训练的代码，并发起训练过程，评估训练结果，完成后可以一键保存训练完成的模型数据，为开发人员提供了便捷的一站式开箱即用的开发环境。

## 2.2 Notebook创建接口

### 2.2.1 简要描述

创建 Notebook 实例。

### 2.2.2 接口约束

无。

### 2.2.3 URL 格式

`https://{endpoint}/api/aidp/v1/notebook_modelview`

URL 参数说明：

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为aidp服务
/v1	必须	string	表示接口版本号为：v1
/notebook_modelview	必须	string	表示为notebook资源

### 2.2.4 请求方式

POST。

### 2.2.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

请求 body 参数:

参数	参数类型	是否必选	描述	示例
project	int	是	项目ID	36
name	string	是	Notebook实例名称	useradmin-3dd3
describe	string	是	Notebook实例描述	1841132031866568705-notebook
imageType	string	是	镜像类型	preset
imageGpuType	string	是	GPU类型	nvidia
images	string	是	镜像地址	image.ccos.io/iaas_pub/notebook:vscode-gpu
specType	string	是	规格类型	compute
resourceId	string	是	资源ID	AIDP-ffa8f37b95a24b439de3966499a1b3e6
resource_memory	string	是	内存规格(G)	100
resource_cpu	string	是	CPU规格(核)	16
resource_gpu	string	是	GPU规格(型号和数量)	1(H800)
vpcEnabled	boolean	是	是否启用VPC	true
vpc_id	string	否	VPC ID (如果启用VPC)	vpc-uavyo483tfgf0c4wfd
subnet_id	string	否	子网ID (如果启用VPC)	snet-v11escbrbofoxc8t6
eip_id	string	否	弹性公网IP ID (如果有)	eip-y0yps5bb5r1hjwrs1
expand	string	是	JSON格式的扩展参数	{}

## 2.2.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	
- code	string	用户id
- message	string	用户名
- result	object	Notebook相关信息
- status	string	状态码

## 2.2.7 示例

请求示例:

```
POST 'https://{endpoint}/api/aidp/v1/notebook '  
--header 'Content-Type: application/json'  
--data-raw      '{"project":36,"name":"useradmin-3dd3","describe":"1841132031866568705-  
notebook","imageType":"preset","imageGpuType":"nvidia","images":"image.ccos.io/iaas_pub/not  
ebook:vscode-gpu","specType":"compute","resourceId":"AIDP-  
ffa8f37b95a24b439de3966499a1b3e6","resource_memory":"100","resource_cpu":"16","resource  
_gpu":"1(H800)","vpcEnabled":true,"vpc_id":"vpc-uavyo483tfgf0c4wfd","subnet_id":"snet-  
v11escbrbofoxc8t6","eip_id":"eip-y0yps5bb5r1hjwrs1","expand":"{}","specCode":"calculate"} '
```

正常返回示例:

```
HTTP/1.1 200 OK  
Content-Type:application/json  
{  
  "requestId": "e67c984c6fdc4c678c8dd2f382bec18a",  
  "code": "Success",  
  "message": "成功",  
  "data": {  
    "code": "",  
    "message": "success",  
    "result": {  
      "describe": "1841132031866568705-notebook",  
      "eip_id": "",
```

```

    "expand": {"ingress_domain": "doko-svc-snet-
qnh9vzdblvyjxz1ux.jupyter", "is_v5": "true", "port_num": "89", "portpool_id": "portPool-
bbese0w1b69wrx5ywwq", "public_ccr_domain": "cke-public-
ccr.registry.region1.cecstackzsarm.cn", "public_ccr_internal_eip": "100.94.11.129", "user_name
": "user@admin"},
    "id": 44,
    "images": "image.ccos.io/iaas_pub/notebook:vscode-gpu",
    "name": "useradmin-3dd3",
    "project": 36,
    "resource_cpu": "16",
    "resource_gpu": "1(H800)",
    "resource_memory": "100G",
    "subnet_id": "",
    "volume_mount": "kubeflow-archives(pvc):/archives,kubeflow-user-workspace(pvc):/mnt",
    "vpc_id": "",
    "working_dir": ""
  },
  "status": 0
}
}

```

## 2.3 Notebook编辑接口

### 2.3.1 简要描述

编辑 Notebook 实例。

### 2.3.2 接口约束

无。

### 2.3.3 URL 格式

[https://{endpoint}/api/aidp/v1/notebook\\_modelview/api/{notebookId}](https://{endpoint}/api/aidp/v1/notebook_modelview/api/{notebookId})

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式

参数	是否必选	参数类型	描述
/aidp	必须	string	表示为aidp服务
/v1	必须	string	表示接口版本号为：v1
/notebook_modelview	必须	string	表示为notebook资源
{notebookId}	必须	Int	表示开发机ID

## 2.3.4 请求方式

PUT。

## 2.3.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

请求 body 参数：

参数	参数类型	是否必选	描述	示例
name	string	是	Notebook实例名称	useradmin-3dd3
describe	string	是	Notebook实例描述	1841132031866568705-notebook-new
specType	string	是	规格类型	compute
resourceId	string	是	资源ID	AIDP-c01bb1977eaf4f4dbe0756cfcaa8d7f2
imageType	string	是	镜像类型	preset
images	string	是	镜像地址	image.ccos.io/iaas_pub/notebook:vscode-cpu
resource_memory	string	是	内存规格 (G)	4G
resource_cpu	string	是	CPU规格 (核)	2
resource_gpu	string	是	GPU规格 (型号和数量)	0
expand	string	是	JSON格式的扩展参数	{具体扩展参数}

## 2.3.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	
- code	string	用户id
- message	string	用户名
- result	object	Notebook相关信息
- status	string	状态码

## 2.3.7 示例

请求示例:

```
PUT 'https://{endpoint}/api/aidp/v1/notebook/api/44'  
--header 'Content-Type: application/json'  
--data-raw      {"name":"useradmin-3dd3","describe":"1841132031866568705-notebook-  
new","specType":"compute","resourceId":"AIDP-  
c01bb1977eaf4f4dbe0756cfcaa8d7f2","imageType":"preset","images":"image.ccos.io/iaas_pub/n  
otebook:vscode-  
cpu","resource_memory":"4","resource_cpu":"2","resource_gpu":"0","expand":{"ingress_domain  
":"doko-svc-snet-  
qnh9vzdblvvyjxz1ux.jupyter","is_v5":"true","port_num":"89","portpool_id":"portPool-  
bbese0w1b69wrx5ywwq","public_ccr_domain":"cke-public-  
ccr.registry.region1.cecstackzsarm.cn","public_ccr_internal_eip":"100.94.11.129","user_name  
":"user@admin"},"specCode":"standardCpu"}
```

正常返回示例:

```
HTTP/1.1 200 OK  
Content-Type: application/json  
{  
  "requestId": "19b52c99a9ad4280bf051c47d78da77c",  
  "code": "Success",  
  "message": "成功",  
  "data": {  
    "code": "",
```

```

    "message": "success",
    "result": {
        // 响应结果详情
    },
    "status": 0
  }
}

```

## 2.4 停止Notebook接口

### 2.4.1 简要描述

停止 Notebook 实例。

### 2.4.2 接口约束

无。

### 2.4.3 URL 格式

https://{endpoint}/api/aidp/v1/resource/stopInstance

URL 参数说明：

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为aidp服务
/v1	必须	string	表示接口版本号为：v1
/resource	必须	string	表示为操作resource
stopInstance	必须	string	表示停止开发机

### 2.4.4 请求方式

GET。

### 2.4.5 请求参数

请求 header 参数：



参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

请求 param 参数:

参数	参数类型	是否必选	描述	示例
instanceId	string	是	实例的ID	44

## 2.4.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	boolean	操作结果，true为成功

## 2.4.7 示例

请求示例:

```
GET 'https://{endpoint}/api/aidp/v1/resource/stopInstance?instanceId=44'
--header 'Content-Type: application/json'
```

正常返回示例:

HTTP/1.1 200 OK

Content-Type:application/json

```
{
  "requestId": "35944b25fcb2439a9312d09fb0cef304",
  "code": "Success",
  "message": "成功",
  "data": true
}
```

## 2.5 重置Notebook接口

### 2.5.1 简要描述

重置 Notebook 实例。

### 2.5.2 接口约束

无。

### 2.5.3 URL 格式

https://{endpoint} /api/aidp/alpha/notebook\_modelview/api/reset/restful/{instanceId}

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/alpha	必须	string	表示接口版本号为: alpha
/notebook_modelview	必须	string	表示为操作rnotebook
{instanceId}	必须	Int	表示开发机ID

### 2.5.4 请求方式

GET。

### 2.5.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型, 根据实际情况传参

请求 path 参数:

参数	参数类型	是否必选	描述	示例
instanceId	string	是	实例的ID	44

## 2.5.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	string	操作结果

## 2.5.7 示例

请求示例：

```
GET 'https://{endpoint}/api/aidp/alpha/notebook/api/reset/restful/44'  
--header 'Content-Type: application/json'
```

正常返回示例：

```
HTTP/1.1 200 OK  
Content-Type:application/json  
{  
  "requestId": "35944b25fcb2439a9312d09fb0cef304",  
  "code": "Success",  
  "message": "成功",  
  "data": "/frontend/dev/dev_online/notebook"  
}
```

## 2.6 Notebook删除接口

### 2.6.1 简要描述

删除 Notebook 实例。

### 2.6.2 接口约束

无。

### 2.6.3 URL 格式

```
https://{endpoint} /api/aidp/v1/deleteInstance
```

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/v1	必须	string	表示接口版本号为: v1
/deleteInstance	必须	string	表示为删除实例操作

## 2.6.4 请求方式

DELETE。

## 2.6.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型, 根据实际情况传参

请求 param 参数:

参数	参数类型	是否必选	描述	示例
id	int	是	实例ID	44
type	int	是	删除类型	1

## 2.6.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	
- code	string	用户id
- message	string	用户名
- result	object	Notebook相关信息
- status	string	状态码

## 2.6.7 示例

请求示例:

```
DELETE 'https://{endpoint}/api/aidp/v1/deleteInstance?id=44&type=1 '  
--header 'Content-Type: application/json'
```

正常返回示例:

```
HTTP/1.1 200 OK
```

```
Content-Type: application/json
```

```
{  
  "requestId": "e67c984c6fdc4c678c8dd2f382bec18a",  
  "code": "Success",  
  "message": "成功",  
  "data": {  
    "code": "",  
    "message": "success",  
    "result": {},  
    "status": 0  
  }  
}
```

# 3 在线服务 API 参考

## 3.1 产品功能简介

在线服务是将训练好的机器学习模型用于进行预测的过程，是实现机器学习系统价值的关键一步。在模型训练完成后,我们需要将模型部署到生产环境中，接收输入数据并返回预测结果。一个好的模型推理系统应具备高效的推理速度、低延迟的请求响应、自动扩缩容的部署架构、丰富的监控指标、多种硬件支持等。实际部署中，可能需要处理大量用户请求，此时我们要考虑请求队列、批处理、缓存等机制来优化吞吐量。并且要确保推理系统的稳定性和可靠性，避免单点故障。通过建立健壮的模型推理平台，通过完善的模型推理功能，可以将机器学习模型轻松转化为可靠、可扩展的在线服务，将机器学习从实验研究转化为可产生业务价值的系统，使得机器学习算法真正造福于客户业务。

## 3.2 在线服务创建接口

### 3.2.1 简要描述

创建在线服务。

### 3.2.2 接口约束

无。

### 3.2.3 URL 格式

`https://{endpoint}/api/aidp/v1/inferenceservice_modelview/api/`

URL 参数说明：

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/v1	必须	string	表示接口版本号为：v1
/inferenceservice_modelview	必须	string	表示为在线服务

### 3.2.4 请求方式

POST。

### 3.2.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型, 根据实际情况传参

请求 body 参数:

参数	参数类型	是否必选	描述	示例
project	int	是	项目ID	36
service_name	string	是	推理服务名称	serving-useradmin-8c55-new
label	string	是	服务标签	abc
describe	string	是	推理服务描述	我的推理服务
infer_backend	string	是	推理后端类型	huggingface
specType	string	是	规格类型	compute
resourceId	string	是	资源ID	AIDP-ffa8f37b95a24b439de3966499a1b3e6
vpc_id	string	是	VPC ID (如果启用VPC)	vpc-uavyo483tfgf0c4wfd
subnet_id	string	是	子网ID (如果启用VPC)	snet-v11escbrbofoxc8t6
eip_id	string	是	弹性公网IP ID	eip-mzo2wce2tu9ouqdfad
min_replicas	int	是	最小副本数	1
max_replicas	int	是	最大副本数	1
hpa	string	是	水平自动扩缩策略 (HPA) 参数	cpu:50%,gpu:50%
priority	int	是	服务优先级	1
eip	string	是	服务绑定的弹性公网IP	10.252.128.161
library_option	int	是	库选项	0
service_type	string	是	服务类型	serving
model_source	int	是	模型来源	0
env	string	否	环境变量	-
model_id	int	是	模型ID	1

参数	参数类型	是否必选	描述	示例
image_source	int	是	镜像来源	0
quantization_bit	int	是	量化位	4

### 3.2.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	

### 3.2.7 示例

请求示例：

```
POST 'https://{endpoint} /api/aidp/v1/inferenceservice_modelview/api/ '
--header 'Content-Type: application/json'
--data-raw '{"project":36,"service_name":"serving-useradmin-8c55-new","label":"abc","describe":"
我的推理服务","infer_backend":"huggingface","specType":"compute","resourceId":"AIDP-
ffa8f37b95a24b439de3966499a1b3e6","vpc_id":"vpc-uavyo483tfgf0c4wfd","subnet_id":"snet-
v11escbbrbofoxc8t6","eip_id":"eip-
mzo2wce2tu9ouqdfad","min_replicas":1,"max_replicas":1,"hpa":"cpu:50%,gpu:50%","priority":1,"
eip":"10.252.128.161","library_option":0,"service_type":"serving","model_source":0,"env":"","mod
el_id":1,"image_source":0,"quantization_bit":4}'
```

正常返回示例：

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "requestId": "e67c984c6fdc4c678c8dd2f382bec18a",
  "code": "Success",
  "message": "成功",
  "data": {}
}
```



## 3.3 在线服务编辑接口

### 3.3.1 简要描述

编辑在线服务。

### 3.3.2 接口约束

无。

### 3.3.3 URL 格式

https://{endpoint}/api/aidp/v1/inferenceservice\_modelview/api/{serviceId}

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为aidp服务
/v1	必须	string	表示接口版本号为: v1
/inferenceservice_modelview	必须	string	表示为在线服务
{serviceId}	必须	Int	表示为在线服务ID

### 3.3.4 请求方式

PUT。

### 3.3.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型, 根据实际情况传参

请求 body 参数:

参数	参数类型	是否必选	描述	示例
project	int	是	项目ID	36
service_name	string	是	推理服务名称	serving-useradmin-8c55-new-1
label	string	是	服务标签	abc-1

参数	参数类型	是否必选	描述	示例
describe	string	是	推理服务描述	我的推理服务-1
model_name	string	是	模型名称	cv-tinytas-object-detection-damoyolo
model_path	string	是	模型路径	/mnt/workspace/.cache/models cope
images	string	是	镜像地址	image.cestc.cn/iaas_pub/mod elscope:base-cuda11.8- python3.8
command	string	是	容器启动命令	/src/docker/entrypoint.sh python app.py web
ports	string	是	容器暴露的端口	80
infer_backend	string	是	推理后端类型	huggingface
specType	string	是	规格类型	compute
resourceId	string	是	资源ID	AIDP- ffa8f37b95a24b439de3966499 a1b3e6
vpc_id	string	是	VPC ID（如果启用VPC）	vpc-uavyo483tfgf0c4wfd
subnet_id	string	是	子网ID（如果启用VPC）	snet-v11escbrbofoxc8t6
eip_id	string	是	弹性公网IP ID	eip-mzo2wce2tu9ouqdfad
min_replicas	int	是	最小副本数	1
max_replicas	int	是	最大副本数	1
hpa	string	是	水平自动扩缩策略 （HPA）参数	cpu:50%,gpu:50%
priority	int	是	服务优先级	1
eip	string	是	服务绑定的弹性公网IP	10.252.128.161
business_type	string	是	业务类型	default
library_option	int	是	库选项	0
service_type	string	是	服务类型	serving
model_source	int	是	模型来源	0
env	string	否	环境变量配置	-
model_id	int	是	模型ID	1
image_source	int	是	镜像来源	0
quantization_bit	int	是	量化位	8

### 3.3.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	

### 3.3.7 示例

请求示例:

```
PUT 'https://{endpoint} /api/aidp/v1/inferenceservice_modelview/api/'
--header 'Content-Type: application/json'
--data-raw      '{"project":36,"service_name":"serving-useradmin-8c55-new-1","label":"abc-1","describe":"我的推理服务-1","model_name":"cv-tinynas-object-detection-damoyolo","model_path":"/mnt/workspace/.cache/modelscope","images":"image.cestc.cn/iaas_public/modelscope:base-cuda11.8-python3.8","command":"/src/docker/entrypoint.sh python app.py web","ports":"80","infer_backend":"huggingface","placeholder":"","specType":"compute","resourceId":"AIDP-ffa8f37b95a24b439de3966499a1b3e6","vpc_id":"vpc-uavyo483tfgf0c4wfd","subnet_id":"snet-v11escbrbofoxc8t6","eip_id":"eip-mzo2wce2tu9ouqdfad","min_replicas":1,"max_replicas":1,"hpa":"cpu:50%,gpu:50%","priority":1,"eip":"10.252.128.161","business_type":"default","library_option":0,"library_config":"","service_type":"serving","model_source":0,"env":"APPNAME=cv-tinynas-object-detection-damoyolo\nAIHUB_MODEL_CONFIG_PATH=/config.json","model_id":1,"image_source":0,"quantization_bit":8}'
```

正常返回示例:

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "requestId": "1f826a0875c14be493349f15e7bc35ad",
  "code": "Success",
  "message": "成功",
  "data": {
    "code": "",
    "message": "success",
    "result": {
      // 根据上表填写的属性
```

```

    },
    "status": 0
  }
}

```

## 3.4 在线服务发布接口

### 3.4.1 简要描述

发布在线服务。

### 3.4.2 接口约束

无。

### 3.4.3 URL 格式

https://{endpoint}/api/aidp/alpha/inferenceservice\_modelview/deploy/prod/restful/{serviceld}

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为aidp服务
/v1	必须	string	表示接口版本号为: v1
/inferenceservice_modelview	必须	string	表示为在线服务
{serviceld}	必须	Int	表示为在线服务ID

### 3.4.4 请求方式

GET。

### 3.4.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型, 根据实际情况传参

请求 body 参数:

该接口可能不需要请求体参数, 操作通过 URL 中指定的服务 ID 进行。

### 3.4.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	

### 3.4.7 示例

请求示例:

```
GET 'https://{endpoint}/api/aidp/alpha/inferenceservice_modelview/deploy/prod/restful/1'  
--header 'Content-Type: application/json'
```

正常返回示例:

```
HTTP/1.1 200 OK
```

```
Content-Type:application/json
```

```
{  
  "requestId": "1f826a0875c14be493349f15e7bc35ad",  
  "code": "Success",  
  "data": "/frontend/service/inferenceservice/inferenceservice_manager",  
  "message": "Success"  
}
```

## 3.5 在线服务下线接口

### 3.5.1 简要描述

下线在线服务。

### 3.5.2 接口约束

无。

### 3.5.3 URL 格式

https://{endpoint}/api/aidp/alpha/inferenceservice\_modelview/clear/restful/{serviceId}

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/v1	必须	string	表示接口版本号为: v1
/inferenceservice_modelview	必须	string	表示为在线服务
{serviceId}	必须	Int	表示为在线服务ID

### 3.5.4 请求方式

GET。

### 3.5.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型, 根据实际情况传参

请求 body 参数:

该接口可能不需要请求体参数, 操作通过 URL 中指定的服务 ID 进行。

### 3.5.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	

### 3.5.7 示例

请求示例:

```
GET 'https://{endpoint}/api/aidp/alpha/inferenceservice_modelview/clear/restful/1'  
--header 'Content-Type: application/json'
```

正常返回示例:

HTTP/1.1 200 OK

Content-Type:application/json

```
{  
  "requestId": "1f826a0875c14be493349f15e7bc35ad",  
  "code": "Success",  
  "data": "/frontend/service/inferenceservice_manager",  
  "message": "Success"  
}
```

## 3.6 在线服务调试接口

### 3.6.1 简要描述

调试在线服务，返回调试接口的 URL。

### 3.6.2 接口约束

无。

### 3.6.3 URL 格式

https://{endpoint}/api/aidp/alpha/inferenceservice\_modelview/deploy/debug/restful/{serviceId}

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/v1	必须	string	表示接口版本号为: v1
/inferenceservice_modelview	必须	string	表示为在线服务
{serviceId}	必须	Int	表示为在线服务ID
requestInfo	必须	String	Region ID

### 3.6.4 请求方式

GET。

### 3.6.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

请求 body 参数：

该接口可能不需要请求体参数，操作通过 URL 中指定的服务 ID 进行。

### 3.6.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	

### 3.6.7 示例

请求示例：

GET

```
https://cecstackzsarm.user.cecloud.com/api/aidp/alpha/inferenceservice_modelview/deploy/debug/restful/123?requestInfo=region1
```

```
--header 'Content-Type: application/json'
```

正常返回示例：

HTTP/1.1 200 OK

Content-Type:application/json

```
{  
  "code": "Success",  
  "data": "",  
  "message": "调试任务资源准备中，等待服务状态变为调试中，再次点击调试按钮即可进入调试页面"  
}
```



## 3.7 在线服务清理接口

### 3.7.1 简要描述

清理在线服务。

### 3.7.2 接口约束

无。

### 3.7.3 URL 格式

https://{endpoint}/api/aidp/v1/deleteInstance

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/v1	必须	string	表示接口版本号为: v1
/deleteInstance	必须	string	表示为删除实例

### 3.7.4 请求方式

DELETE。

### 3.7.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型, 根据实际情况传参

请求 body 参数:

参数	参数类型	是否必选	描述	示例
id	int	是	实例ID	169
type	int	是	删除类型	3

### 3.7.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	
- code	string	用户id
- message	string	用户名
- result	object	在线服务相关信息
- status	string	状态码

### 3.7.7 示例

请求示例:

```
DELETE 'https://{endpoint}/api/aidp/v1/deleteInstance?id=169&type=3'  
--header 'Content-Type: application/json'
```

正常返回示例:

```
HTTP/1.1 200 OK  
Content-Type:application/json  
{  
  "requestId": "c18c2b178e91494caeb01c781f05a6a7",  
  "code": "Success",  
  "message": "成功",  
  "data": {  
    "code": "",  
    "message": "success",  
    "result": {  
      // 根据实例的具体信息，列出所有相关属性  
    },  
    "status": 0  
  }  
}
```

# 4 公共资源池 API 参考

## 4.1 产品功能简介

GPU 公共资源池是一个为多个用户提供共享 GPU 计算资源的服务。这种资源池允许用户根据需求动态分配和释放 GPU 资源，优化资源利用率，并支持各种计算密集型任务，如深度学习训练、科学计算和图形渲染等。

## 4.2 购买资源包接口

### 4.2.1 简要描述

购买资源包。

### 4.2.2 接口约束

无。

### 4.2.3 URL 格式

`https://{endpoint}/api/aidp/v1/resource`

URL 参数说明：

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/v1	必须	string	表示接口版本号为：v1
/resource	必须	string	表示为资源类型的操作

### 4.2.4 请求方式

POST。

### 4.2.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

请求 body 参数:

参数	参数类型	是否必选	描述	示例
queryPriceResult	object	是	价格查询结果详情	-
- showPriceSwitch	string	是	是否展示价格	"on"
- chargeSwitch	string	是	是否开启计费	"off"
- totalPrice	object	是	总价信息	-
- - price	int	是	价格（单位：分）	7200000
- - discount	int	是	折扣（单位：分）	0
- - total	int	是	总计（单位：分）	7200000
- - displayPrice	int	是	展示价格（单位：元）	72000
- - priceDetailDTOS	array	是	价格细节列表	-
- unitPrice	object	是	单价信息	-
- - unitPriceDetailDTOS	array	是	单价细节列表	-
- tips	string	是	提示信息	"--"
ratingKey	string	是	计费关键字	"ByHours"
specCode	string	是	规格代码	"standardCpu"
cycleCount	int	是	周期数量	1
cycleType	string	是	周期类型（如：月、年）	"Month"
amount	int	是	数量	1
tenantId	string	是	租户ID	"1841781662849977753"
departmentId	string	是	部门ID	"1841781662849977754"
name	string	是	名称	"useradmin-bg11"
azId	string	是	可用区ID	"AZ1"
billingMode	string	是	计费模式	"ByCycle"
bizType	string	是	业务类型（如：创建、更新）	"Create"
regionId	string	是	区域ID	"region1"

## 4.2.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	string	返回的数据

## 4.2.7 示例

请求示例:

```
POST 'https://{endpoint}/api/aidp/v1/resource'  
--header 'Content-Type: application/json'  
--data-raw  
{  
  "queryPriceResult": {"showPriceSwitch": "on", "chargeSwitch": "off", "totalPrice": {"price": 7200000, "discount": 0, "total": 7200000, "displayPrice": 72000, "priceDetailDTOS": [{"serviceCode": "AIDP", "serviceDisplayName": "智算产品", "specCode": "standardCpu", "specName": "2c | 4g | 0 卡通用型CPU(2c4g)", "instanceId": "", "bizType": "Create", "price": 7200000, "discount": 0, "total": 7200000, "ratingKeyPriceDTOS": [{"pricelId": 10161, "ratingKey": "ByHours", "count": 100, "fee": 7200000, "feeUnit": "分", "ratingVal": 100, "feeMeasure": "10403", "cycleType": "Month", "priceType": "ByCycleDefault"}]}]}, "unitPrice": {"unitPriceDetailDTOS": [{"serviceCode": "AIDP", "serviceDisplayName": "智算产品", "specCode": "standardCpu", "priceDescDetailDTOS": [{"pricelId": 10161, "priceName": "按CPU核数计费", "fee": "7200000", "feeUnit": "分", "feeMeasureId": "10403", "cycleType": "月", "priceType": "ByCycleDefault", "ratingKey": "ByHours", "ratingVal": 100, "ratingkeyName": "按CPU核数", "resMeasureId": "Times", "resUnit": "1", "unitPriceDesc": "720 元/次/月", "display": "72000/月"}]}]}, "tips": "--  
"}, {"ratingKey": "ByHours", "specCode": "standardCpu", "cycleCount": 1, "cycleType": "Month", "amount": 1, "tenantId": "1841781662849977753", "departmentId": "1841781662849977754", "name": "useradmin-bg11", "azId": "AZ1", "billingMode": "ByCycle", "bizType": "Create", "regionId": "region1"} ]
```

正常返回示例:

```
HTTP/1.1 200 OK  
Content-Type: application/json  
{  
  "requestId": "a9cf9be0cefd49c887ded101da1452cb",  
  "code": "Success",  
  "message": "成功",  
  "data": "221189662184701952"
```

}

## 4.3 资源包列表接口

### 4.3.1 简要描述

查询资源包列表。

### 4.3.2 接口约束

无。

### 4.3.3 URL 格式

https://{endpoint}/api/aidp/v1/resource/list

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/v1	必须	string	表示接口版本号为: v1
/resource	必须	string	表示为资源类型的操作
/list	必须	string	表示为列表查询

### 4.3.4 请求方式

POST。

### 4.3.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型, 根据实际情况传参

请求 param 参数:

参数	参数类型	是否必选	描述	示例
requestInfo	string	是	用于筛选的额外信息，如区域ID	requestInfo

请求 body 参数：

参数	参数类型	是否必选	描述	示例
pageSize	int	是	每页显示的资源数量	pageSize
pageNum	int	是	当前页码	pageNum
specType	string	是	资源类型，如 "compute"	specType

### 4.3.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	string	返回的数据

data 对象属性：

参数	参数类型	描述
size	int	每页大小
current	int	当前页码
total	int	总资源数量
totalPage	int	总页数
list	array	资源列表

list 中的对象属性：

参数	参数类型	描述
id	int	资源ID
name	string	资源名称
orderId	string	订单ID
tenantId	string	租户ID
departmentId	string	部门ID

参数	参数类型	描述
project	string	项目名称
specCode	string	规格代码
specName	string	规格名称
specCPU	string	CPU规格
specGpu	string	GPU规格
specMemory	string	内存规格
restCPU	string	剩余CPU规格
restGpu	string	剩余GPU规格
restMemory	string	剩余内存规格
endTime	string	结束时间
createUser	string	创建用户
createTime	int	创建时间（时间戳）
updateUser	string	更新用户
updateTime	int	更新时间（时间戳）
resourceType	string	资源类型
resourceState	int	资源状态
regionId	string	区域ID
azId	string	可用区ID
cellId	string	单元格ID
resourceGroupId	string	资源组ID
billingMode	string	计费模式
perPrice	int	单价（单位：分）
feeUnit	int	费用单位
cycleType	string	计费周期类型
delete	int	是否已删除
projectName	string	项目名称
tenantName	string	租户名称
departmentName	string	部门名称
resourceGroupName	string	资源组名称
resourceStateName	string	资源状态名称
totalAmount	int	总数量
usedAmount	int	已使用数量



参数	参数类型	描述
restAmount	int	剩余数量
expireTime	string	到期时间
instanceType	string	实例类型
durationNum	int	持续时间数量
beforeAmount	int	修改前数量
afterAmount	int	修改后数量
cycleCount	int	计费周期数量
specStorage	string	存储规格
storageSize	string	存储大小
usageType	string	使用类型
busId	string	业务ID
instanceName	string	实例名称
instanceTypeCode	int	实例类型代码
internalGpuType	string	内部GPU类型

### 4.3.7 示例

请求示例:

```
POST 'https://{endpoint}/api/aidp/v1/resource/list?requestInfo=region1'
--header 'Content-Type: application/json'
--data-raw '{"pageSize":10,"pageNum":1,"specType":"compute"}
```

正常返回示例:

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "requestId": "",
  "code": "Success",
  "message": "成功",
  "data": {
    "size": 10,
    "current": 1,
    "total": 56,
    "totalPage": 6,
    "list": [
```

```

    {...}
  ]
}
}

```

## 4.4 资源包详情接口

### 4.4.1 简要描述

查询资源包详情。

### 4.4.2 接口约束

无。

### 4.4.3 URL 格式

https://{endpoint}/api/aidp/v1/resource/detail/{resourceId}

URL 参数说明：

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为aidp服务
/v1	必须	string	表示接口版本号为：v1
/resource	必须	string	表示为资源类型的操作
/detail	必须	string	表示为详情查询
{resourceId}	必须	string	表示为资源包ID

### 4.4.4 请求方式

GET。

### 4.4.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

请求 body 参数:

该接口不需要请求参数, 资源包 ID 通过 URL 传递。

## 4.4.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	string	返回的数据

data 对象属性:

参数	参数类型	描述
id	int	资源ID
name	string	资源名称
orderId	string	订单ID
tenantId	string	租户ID
departmentId	string	部门ID
project	string	项目名称
specCode	string	规格代码
specName	string	规格名称
specCPU	string	CPU规格
specGpu	string	GPU规格
specMemory	string	内存规格
restCPU	string	剩余CPU规格
restGpu	string	剩余GPU规格
restMemory	string	剩余内存规格
endTime	string	结束时间
createUser	string	创建用户
createTime	int	创建时间 (时间戳)
updateUser	string	更新用户
updateTime	int	更新时间 (时间戳)
resourceType	string	资源类型
resourceState	int	资源状态
regionId	string	区域ID

参数	参数类型	描述
azId	string	可用区ID
cellId	string	单元格ID
resourceGroupId	string	资源组ID
billingMode	string	计费模式
perPrice	int	单价（单位：分）
feeUnit	int	费用单位
cycleType	string	计费周期类型
delete	int	是否已删除
projectName	string	项目名称
tenantName	string	租户名称
departmentName	string	部门名称
resourceGroupName	string	资源组名称
resourceStateName	string	资源状态名称
totalAmount	int	总数量
usedAmount	int	已使用数量
restAmount	int	剩余数量
expireTime	string	到期时间
instanceType	string	实例类型
durationNum	int	持续时间数量
beforeAmount	int	修改前数量
afterAmount	int	修改后数量
cycleCount	int	计费周期数量
specStorage	string	存储规格
storageSize	string	存储大小
usageType	string	使用类型
busId	string	业务ID
instanceName	string	实例名称
instanceTypeCode	int	实例类型代码
internalGpuType	string	内部GPU类型

#### 4.4.7 示例

请求示例：

```
GET 'https://{endpoint}/api/aidp/v1/resource/detail/AIDP-08ba092ac675457fb4ecf9e6a2ae97b4'  
--header 'Content-Type: application/json'
```

正常返回示例:

HTTP/1.1 200 OK

Content-Type:application/json

```
{  
  "requestId": "e3c74faad8ee49b28f396178c847af04",  
  "code": "Success",  
  "message": "成功",  
  "data": {  
    "id": 78,  
    "name": "useradmin-bg11",  
    "orderId": "221189662184701952",  
    "tenantId": "1841781662849977753",  
    "departmentId": "1841781662849977754",  
    "project": "",  
    "specCode": "standardCpu",  
    "specName": "通用型 CPU 无 GPU(2c4g)|standardCpu",  
    "specCPU": "2",  
    "specGpu": "0",  
    "specMemory": "4",  
    "restCPU": "",  
    "restGpu": "",  
    "restMemory": "",  
    "endTime": "2025-07-11 22:29:12",  
    "createUser": "",  
    "createTime": 1720708152150,  
    "updateUser": "",  
    "updateTime": 1720708152150,  
    "resourceType": "compute",  
    "resourceState": 1,  
    "regionId": "region1",  
    "azId": "AZ1",  
    "cellId": "",  
    "resourceId": "AIDP-08ba092ac675457fb4ecf9e6a2ae97b4",  
    "resourceGroupId": "",  
    "billingMode": "ByCycle",
```

```
"perPrice": 7200000,
"feeUnit": 0,
"cycleType": "Month",
"delete": 0,
"projectName": "",
"tenantName": "guodx",
"departmentName": "guodx",
"resourceGroupName": "",
"resourceStateName": "正常",
"totalAmount": 100,
"usedAmount": 0,
"restAmount": 100,
"expireTime": "",
"instanceType": "",
"durationNum": 0,
"beforeAmount": 0,
"afterAmount": 0,
"cycleCount": 0,
"specStorage": "",
"storageSize": "",
"usageType": "2",
"busId": "",
"instanceName": "",
"instanceTypeCode": 0,
"internalGpuType": ""
}
}
```

# 5 存储资源 API 参考

## 5.1 产品功能简介

存储资源是信息技术基础设施中的关键组成部分，用于数据的持久化保存。目前主要提供并行文件存储的能力。并行存储作为一种高性能、高可扩展的分布式存储系统，在各个领域都有着广泛的应用前景。通过深入了解并行存储的架构、工作原理和实际应用场景，我们可以更好地理解这一强大的分布式存储解决方案，并利用它来解决实际业务中的数据存储和处理问题。

## 5.2 存储资源购买接口

### 5.2.1 简要描述

购买存储资源。

### 5.2.2 接口约束

无。

### 5.2.3 URL 格式

`https://{endpoint}/api/aidp/v1/storage/resource`

URL 参数说明：

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/v1	必须	string	表示接口版本号为：v1
/storage/resource	必须	string	表示为存储资源

### 5.2.4 请求方式

POST。

### 5.2.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

请求 body 参数:

参数	参数类型	是否必选	描述	示例
queryPriceResult	object	是	价格查询结果详情	-
- showPriceSwitch	string	是	是否展示价格	"on"
- chargeSwitch	string	是	是否开启计费	"off"
- totalPrice	object	是	总价信息	-
- - price	int	是	价格（单位：分）	10000000
- - discount	int	是	折扣（单位：分）	0
- - total	int	是	总计（单位：分）	10000000
- - displayPrice	int	是	展示价格（单位：元）	100000
- - priceDetailDTOS	array	是	价格细节列表	-
- unitPrice	object	是	单价信息	-
- - unitPriceDetailDTOS	array	是	单价细节列表	-
- tips	string	是	提示信息	"--"
ratingKey	string	是	计费关键字	"ByAmount"
specCode	string	是	规格代码	"aispace.gpfs11"
cycleCount	int	是	周期数量	1
cycleType	string	是	周期类型（如：月、年）	"Month"
amount	int	是	数量	100
tenantId	string	是	租户ID	"1841781662849977753"
departmentId	string	是	部门ID	"1841781662849977754"
name	string	是	名称	"useradmin-7224"
azId	string	是	可用区ID	"AZ1"
billingMode	string	是	计费模式	"ByCycle"
bizType	string	是	业务类型（如：创建、更新）	"Create"
regionId	string	是	区域ID	"region1"



## 5.2.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	

## 5.2.7 示例

请求示例:

```
POST 'https://{endpoint}/api/aidp/v1/storage/resource '  
--header 'Content-Type: application/json'  
--data-raw  
'{"queryPriceResult":{"showPriceSwitch":"on","chargeSwitch":"off","totalPrice":{"price":10000000,"  
discount":0,"total":10000000,"displayPrice":100000,"priceDetailDTOS":[{"serviceCode":"AIDP","s  
erviceDisplayName":"智          算          产          品  
","specCode":"aispace.gpfs11","specName":"aispace.gpfs11","instanceId":"","bizType":"Create","  
price":10000000,"discount":0,"total":10000000,"ratingKeyPriceDTOS":[{"pricelId":10160,"ratingKe  
y":"ByAmount","count":100,"fee":10000000,"feeUnit":"分  
","ratingVal":100,"feeMeasure":"10403","cycleType":"Month","priceType":"ByCycleDefault"}]}],"un  
itPrice":{"unitPriceDetailDTOS":[{"serviceCode":"AIDP","serviceDisplayName":"智 算 产 品  
","specCode":"aispace.gpfs11","priceDescDetailDTOS":[{"pricelId":10160,"priceName":"按 机 器 规  
格 及 时 长 计 费","fee":"10000000","feeUnit":"分","feeMeasureId":"10403","cycleType":"月  
","priceType":"ByCycleDefault","ratingKey":"ByAmount","ratingVal":100,"ratingkeyName":"按 机 器  
规 格 及 时 长 计 费","resMeasureId":"Times","resUnit":"1","unitPriceDesc":"1000 元/次/月  
","display":"100000/月"}]}],"tips":"--  
"},"ratingKey":"ByAmount","specCode":"aispace.gpfs11","cycleCount":1,"cycleType":"Month","am  
ount":100,"tenantId":"1841781662849977753","departmentId":"1841781662849977754","name":  
"useradmin-7224","azId":"AZ1","billingMode":"ByCycle","bizType":"Create","regionId":"region1"}'
```

正常返回示例:

```
HTTP/1.1 200 OK  
Content-Type:application/json  
{  
  "requestId": "5a0df555c99c43128c88f5fe545a7634",  
  "code": "Success",  
  "message": "成功",
```

```
"data": "221191025148952576"
}
```

## 5.3 存储资源列表接口

### 5.3.1 简要描述

查询存储资源列表。

### 5.3.2 接口约束

无。

### 5.3.3 URL 格式

`https://{endpoint}/api/aidp/v1/storage/resource/list`

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/v1	必须	string	表示接口版本号为: v1
/storage/resource	必须	string	表示为存储资源
/list	必须	string	表示为列表查询

### 5.3.4 请求方式

POST。

### 5.3.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型, 根据实际情况传参

请求 param 参数:

参数	参数类型	是否必选	描述	示例
requestInfo	string	是	筛选条件，例如区域ID	requestInfo

请求 body 参数：

参数	参数类型	是否必选	描述	示例
pageSize	int	是	每页显示的资源数量	pageSize
pageNum	int	是	当前页码	pageNum
name	string	否	资源名称	name
orderId	string	否	订单ID	orderId

### 5.3.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	

data 对象属性：

参数	参数类型	描述
size	int	每页大小
current	int	当前页码
total	int	总资源数量
totalPage	int	总页数
list	array	资源列表

List 数组元素属性：

参数	参数类型	描述
id	int	资源ID
name	string	资源名称
orderId	string	订单ID
tenantId	string	租户ID
departmentId	string	部门ID

参数	参数类型	描述
project	string	项目名称
specCode	string	规格代码
specName	string	规格名称
specCPU	string	CPU规格
specGpu	string	GPU规格
specMemory	string	内存规格
restCPU	string	剩余CPU规格
restGpu	string	剩余GPU规格
restMemory	string	剩余内存规格
endTime	string	结束时间
createUser	string	创建用户
createTime	int	创建时间（时间戳）
updateUser	string	更新用户
updateTime	int	更新时间（时间戳）
resourceType	string	资源类型
resourceState	int	资源状态
regionId	string	区域ID
azId	string	可用区ID
cellId	string	单元格ID
resourceId	string	资源ID
resourceGroupId	string	资源组ID
billingMode	string	计费模式
perPrice	int	单价（单位：分）
feeUnit	int	费用单位
cycleType	string	计费周期类型
delete	int	是否已删除
projectName	string	项目名称
tenantName	string	租户名称
departmentName	string	部门名称
resourceGroupName	string	资源组名称
resourceStateName	string	资源状态名称
totalAmount	int	总数量

参数	参数类型	描述
usedAmount	int	已使用数量
restAmount	int	剩余数量
expireTime	string	到期时间
instanceType	string	实例类型
durationNum	int	持续时间数量
beforeAmount	int	修改前数量
afterAmount	int	修改后数量
cycleCount	int	计费周期数量
specStorage	string	存储规格
storageSize	string	存储大小
usageType	string	使用类型
busId	string	业务ID
instanceName	string	实例名称
instanceTypeCode	int	实例类型代码
internalGpuType	string	内部GPU类型

### 5.3.7 示例

请求示例：

```
POST 'https://{endpoint}/api/aidp/v1/storage/resource '
--header 'Content-Type: application/json'
--data-raw '{"pageSize":10,"pageNum":1,"name":"useradmin-7224"}
```

正常返回示例：

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "requestId": "",
  "code": "Success",
  "message": "成功",
  "data": {
    "size": 10,
    "current": 1,
```

```
"total": 1,  
"totalPage": 1,  
"list": [  
  {  
    // 根据上表填写的属性  
  }  
]  
}  
}
```

# 6 专属资源池 API 参考

## 6.1 产品功能简介

智算专属资源池是一种高性能、可定制的计算环境，专为满足人工智能、深度学习、数据分析等高负载计算任务的需求而设计。它提供专属访问权限，确保用户能够安全、高效地运行其计算工作负载，同时享有高度的灵活性和可扩展性，支持动态资源分配和精细的性能优化。通过集中管理和监控，智算专属资源池帮助用户降低成本、提高生产力，并确保数据处理的安全性和合规性。。

## 6.2 专属资源池购买接口

### 6.2.1 简要描述

购买专属资源池。

### 6.2.2 接口约束

无。

### 6.2.3 URL 格式

`https://{endpoint}/api/aidp/v1/resource`

URL 参数说明：

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/v1	必须	string	表示接口版本号为：v1
/resource	必须	string	表示为专属资源

### 6.2.4 请求方式

POST。

### 6.2.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

请求 body 参数:

参数	是否必选	参数类型	描述	示例
showPriceSwitch	是	string	指示是否显示价格	on
chargeSwitch	是	string	是否启用收费	on
totalPrice	是	object	总价的详细信息。	
price	是	int	原价	
discount	是	int	折扣	
total	是	int	折扣后的总价	
displayPrice	是	int	显示的价格	
priceDetailDTOS	是	array	价格详细信息的数组	
serviceCode	是	string	服务代码	
serviceDisplayName	是	string	服务显示名称	
specCode	是	string	规格代码	
specName	是	string	规格名称	
instanceId	是	string	实例ID	
bizType	是	string	业务类型	
price	是	int	原价	
discount	是	int	折扣	
total	是	int	总价	
ratingKeyPriceDTOS	是	Array	包含计费详细信息的数组	
priceId	是	Number	价格ID	
ratingKey	是	string	计费键	
count	是	int	数量	
fee	是	int	费用	
feeUnit	是	string	费用单位	
ratingVal	是	int	计费值	
feeMeasure	是	string	费用度量单位	
cycleType	是	string	计费周期类型	
priceType	是	string	价格类型	



unitPrice	是	Object	单价详细信息	
unitPriceDetailDTOS	是	array	包含单价详细信息的数组	
serviceCode	是	string	服务代码	
serviceDisplayName	是	string	服务显示名称	
specCode	是	string	规格代码	
priceDescDetailDTOS	是	array	包含价格描述详细信息的数组	
priceId	是	int	价格ID	
priceName	是	string	价格名称	
fee	是	int	费用	
feeUnit	是	string	费用单位	
feeMeasureId	是	string	费用度量单位ID	
cycleType	是	string	计费周期类型	
priceType	是	string	价格类型	
ratingKey	是	string	计费键	
ratingVal	是	string	计费值。	
ratingkeyName	是	string	计费键名称。	
resMeasureId	是	string	资源度量单位ID。	
resUnit	是	string	资源单位。	
unitPriceDesc	是	string	单价描述。	
display	是	string	显示	
tips	是	string	提示信息	
ratingKey	是	string	计费键	
specCode	是	string	规格代码	
cycleCount	是	int	周期计数。	
cycleType	是	string	周期类型。	
amount	是	int	数量。	
tenantId	是	string	租户ID。	
departmentId	是	string	部门ID。	
name	是	string	名称。	
azId	是	string	可用区ID。	
billingMode	是	string	计费模式。	
bizType	是	string	业务类型。	

regionId	是	string	区域ID。	
----------	---	--------	-------	--

## 6.2.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	

## 6.2.7 示例

请求示例：

```
curl 'https://cecstackzsarm.user.cecloud.com/api/aidp/v1/resource' \
-H 'Accept: application/json, text/plain, */*' \
-H 'Connection: keep-alive' \
-H 'Content-Type: application/json' \
-H 'x-auth-validate: true' \
-H 'x-regionid: region1' \
--data-raw
{"queryPriceResult":{"showPriceSwitch":"on","chargeSwitch":"off","totalPrice":{"price":200,"discount":0,"total":200,"displayPrice":2,"priceDetailDTOS":[{"serviceCode":"AIDP","serviceName":"智算产品","specCode":"spec","specName":"spec","instanceId":"","bizType":"Create","price":200,"discount":0,"total":200,"ratingKeyPriceDTOS":[{"priceId":10213,"ratingKey":"ByAmount","count":1,"fee":200,"feeUnit":"分","ratingVal":1,"feeMeasure":"10403","cycleType":"Month","priceType":"ByCycleDefault"}]}],"unitPrice":{"unitPriceDetailDTOS":[{"serviceCode":"AIDP","serviceName":"智算产品","specCode":"spec","priceDescDetailDTOS":[{"priceId":10213,"priceName":"按机器规格及时长计费","fee":"200","feeUnit":"分","feeMeasureId":"10403","cycleType":"月","priceType":"ByCycleDefault","ratingKey":"ByAmount","ratingVal":1,"ratingKeyName":"按机器规格及时长计费","resMeasureId":"Counts","resUnit":"1","unitPriceDesc":"2 元/个/月","display":"2/月"}]}],"tips":"--"}, {"ratingKey":"ByAmount","specCode":"spec","cycleCount":1,"cycleType":"Month","amount":1,"tenantId":"1841781662849977753","departmentId":"1841781662849977754","name":"guodx1-g9f9","azId":"AZ1","billingMode":"ByCycle","bizType":"Create","regionId":"region1"} \
--insecure
```

正常返回示例:

HTTP/1.1 200 OK

Content-Type:application/json

```
{
  "requestId": "eb640ad28e5f4426a77f6e556e4fc062",
  "code": "Success",
  "message": "成功",
  "data": "221357312907018240"
}
```

## 6.3 专属资源池退订接口

### 6.3.1 简要描述

退订资源池。

### 6.3.2 接口约束

无。

### 6.3.3 URL 格式

https://{endpoint}/api/aidp/v1/resource/{resource\_id}

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/v1	必须	string	表示接口版本号为: v1
/resource/{resource_id}	必须	string	表示为专属资源池ID
requestInfo	必须	string	Region id

### 6.3.4 请求方式

DELETE。

### 6.3.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

请求 body 参数:

参数	是否必选	参数类型	描述	示例
queryPriceResult				
showPriceSwitch				
chargeSwitch				
remainFee				
totalFee				
usedFee				
refundDetailDTOS				
serviceCode				
serviceDisplayName				
specCode				
instanceId				
bizType				
refundFee				
refundOrderDTOS				
orderId				
tradeId				
refundFee				
remainFee				
perPrice				
ratingKey				

### 6.3.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	string	订单ID

## 6.3.7 示例

请求示例:

```
curl          'https://cecstackzsarm.user.cecloud.com/api/aidp/v1/resource/AIDP-
f0283cb3f05b469c96ceea0b4476b89e?requestInfo=region1' \
-X 'DELETE' \
-H 'Accept: application/json, text/plain, */*' \
--data-raw
'{"queryPriceResult":{"showPriceSwitch":"on","chargeSwitch":"off","remainFee":194,"totalFee":20
0,"usedFee":6,"refundDetailDTOS":[{"serviceCode":"AIDP","serviceDisplayName":"","specCode":
"spec","instanceId":"AIDP-
f0283cb3f05b469c96ceea0b4476b89e","bizType":"Refund","refundFee":194,"refundOrderDTOS"
:[{"orderId":"221357312907018240","tradeId":"221357312907018241","refundFee":194,"remainF
ee":194}]}}],"perPrice":194,"ratingKey":"ByAmount"}' \
--insecure
```

正常返回示例:

```
Content-Type:application/json
{
  "requestId": "59b06f3fb13249feafd67df8fde9905d",
  "code": "Success",
  "message": "成功",
  "data": "221374549881651200"
}
```

# 7 大模型 API 参考

## 7.1 产品功能简介

大模型服务建立在“模型即服务”(Model-as-a-Service, MaaS)的理念基础之上,围绕数十亿到上万亿参数级的大规模预训练语言模型,通过标准化的 API 提供模型训练、调优、评估等全方位功能与服务,并重点支撑大模型生态,支持开发者使用大模型服务进行语言理解、生成等任务,免去自行训练和部署的高门槛,大幅降低 AI 的使用难度,是 AI 开发平台的重要组成部分,是面向企业开发者的一站式大模型开发及服务运行平台。

## 7.2 大模型预训练、调优(SFT、DPO、RLHF)创建接口

### 7.2.1 简要描述

大模型预训练、调优(SFT、DPO、RLHF)任务创建。

### 7.2.2 接口约束

无。

### 7.2.3 URL 格式

大模型任务类型	URL 格式
预训练	https://{endpoint}/api/aidp/alpha/maas_task_modelview/pt/api?requestInfo={regionId}
调优SFT	https://{endpoint}/api/aidp/alpha/maas_task_modelview/sft/api?requestInfo={regionId}
调优DPO	https://{endpoint}/api/aidp/alpha/maas_task_modelview/dpo/api?requestInfo={regionId}
调优RLHF—奖励模型训练	https://{endpoint}/api/aidp/alpha/maas_task_modelview/rw/api?requestInfo={regionId}
调优RLHF—强化学习训练	https://{endpoint}/api/aidp/alpha/maas_task_modelview/ppo/api?requestInfo={regionId}

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
{regionId}	必须	string	region可用区

## 7.2.4 请求方式

POST。

## 7.2.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

请求 body 参数：

参数	是否必选	参数类型	示例	描述
jobname	是	string	test-0711	作业名称
name	是	string	test-002	任务名称
label	是	string	test-002	任务标签
storage_type	是	int	0	存储类型，0 表示使用默认存储类型
billing_mode	是	string	resourcePool	计费模式
bill_gui	否	string	defaultBill	计费界面，留空使用默认
spec_type	是	string	compute	规格类型，如计算型
node_selector	是	string	train=true	节点选择器，用于指定任务运行的节点
workers	是	int	1	工作节点数量
resource_id	是	string	AIDP-f5adef644a45ca8ab4a56c0f0f63ab	资源ID
framework	是	string	deepspeed	框架类型，如deepspeed
ft_type	是	string	lora	功能类型，如lora
level	是	string	L0	等级
advanced_mode	是	string	close	高级模式，如关闭
framework_template	是	string	stage2	框架模板
aihub_id	是	int	41	AIHub ID
lab_id	否	int	2	关联实验管理ID
maas_task_parent_id	否	null	null	父任务ID，留空表示无父任务
model_path	否	null	null	模型路径，留空表示使用默认

参数	是否必选	参数类型	示例	描述
model_type	否	null	null	模型类型，留空表示使用默认
projectId	是	int	45	项目ID
dataset_ids	是	string	[115]	数据集ID列表
maas	是	int	44	MAAS ID
args	是	string	详细的参数JSON字符串	运行参数
framework_config	是	string	详细的框架配置JSON字符串	框架配置

## 7.2.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	返回的数据对象

data 对象结构

参数	参数类型	描述
code	string	状态码
message	string	状态消息
result	object	任务详细信息
status	int	状态值

result 对象结构:

参数	参数类型	描述
advanced_mode	string	任务是否启用高级模式，例如 "close" 表示关闭高级模式
aihub_id	int	与AIHub相关联的ID
args	string	任务运行参数的JSON字符串
batch_size	int	批处理大小
billing_mode	string	计费模式，例如 "resourcePool"
dataset_id	null	数据集ID，返回中可能为null
dataset_ids	string	数据集ID列表，JSON格式的字符串
default_system	null	默认系统配置，返回中可能为null
epoch	int	训练周期数



参数	参数类型	描述
exclusive	string	独占模式，返回中可能为空字符串
expand	string	扩展配置，返回中可能为空字符串
framework	string	使用的框架，例如 "deepspeed"
framework_config	string	框架配置的JSON字符串
framework_template	string	框架模板名称，例如 "stage2"
ft_type	string	功能类型，例如 "lora"
id	int	任务的唯一标识符
label	string	任务标签
level	string	任务等级，例如 "L0"
maas_task_parent_id	null	父任务ID，返回中可能为null
model_path	string	模型路径，返回中可能为空字符串
model_type	string	模型类型，例如 "llama2_zh"
name	string	任务名称
node_selector	string	节点选择器
output_dir	string	输出目录，返回中可能为空字符串
resource_cpu	string	CPU资源规格，例如 "50"
resource_gpu	string	GPU资源规格，例如 "2(H800)"
resource_id	string	资源ID
resource_memory	string	内存资源规格，例如 "200G"
retry	int	重试次数
spec_type	string	规格类型，例如 "compute"
storage_type	int	存储类型
workers	int	工作节点数量

## 7.2.7 示例

请求示例：

POST

```
'https://{endpoint}/api/aidp/alpha/maas_task_modelview/pt/api/?requestInfo=region1'
```

```
--header 'Content-Type: application/json'
```

```
--data-raw '{ "jobname": "test-0711", "name": "test-002", "label": "test-002", "storage_type": 0, "billing_mode": "resourcePool", "bill_gui": "defaultBill", "spec_type": "compute", "node_selector": "train=true", "workers": 1, "resource_id": "AIDP-f5adeffe644a45ca8ab4a56c0f0f63ab",
```

```

"framework": "deepspeed", "ft_type": "lora", "level": "L0", "advanced_mode": "close", "framework_template": "stage2", "aihub_id": 41, "maas_task_parent_id": null, "model_path": null, "model_type": null, "projectId": 45, "dataset_ids": "[115]", "maas": "44", "args": "{
    \"MAX_LENGTH\": 1024,
    \"LEARNING_RATE\": 0.00005,
    \"LORA_RANK\": 8,
    \"LORA_DROPOUT\": 0.1,
    \"LORA_TARGET\": \"all\",
    \"QUANTIZATION_BIT\": -1,
    \"GRADIENT_CHECKPOINTING\": false,
    \"GRADIENT_ACCUMULATION_STEPS\": 4,
    \"SAVE_STEPS\": 100,
    \"SAVE_TOTAL_LIMIT\": 2,
    \"LOGGING_STEP\": 10,
    \"FLASH_ATTN\": false,
    \"SHIFT_ATTN\": false,
    \"DO_EVAL\": false,
    \"EVALUATION_STRATEGY\": \"\",
    \"EVAL_STEPS\": 10,
    \"EPOCH\": 3,
    \"BATCH_SIZE\": 1,
    \"EVAL_SIZE\": 0.01,
    \"REWARD_MODEL\": \"/.reward\",
    \"PORT\": \"\"
}", "framework_config": "{
    \"train_batch_size\": \"auto\",
    \"train_micro_batch_size_per_gpu\": \"auto\",
    \"gradient_accumulation_steps\": \"auto\",
    \"gradient_clipping\": \"auto\",
    \"zero_allow_untested_optimizer\": true,
    \"fp16\": {
        \"enabled\": \"auto\",
        \"loss_scale\": 0,
        \"initial_scale_power\": 16,
        \"loss_scale_window\": 1000,
        \"hysteresis\": 2,
        \"min_loss_scale\": 1
    },
    \"zero_optimization\": {
        \"stage\": 2,
        \"allgather_partitions\": true,
        \"allgather_bucket_size\": 5e8,
        \"reduce_scatter\": true,
        \"reduce_bucket_size\": 5e8,
        \"overlap_comm\": true,
        \"contiguous_gradients\": true
    }
}"

```

正常返回示例:

HTTP/1.1 200 OK

Content-Type: application/json

```

{
  "requestId": "539ff1f2a578478b996d5b4471e23d7d",
  "code": "Success",
  "message": "成功",
  "data": {
    "code": "",
    "message": "success",
    "result": {
      "advanced_mode": "close",
      "aihub_id": 41,
      "args": "{
        \"MAX_LENGTH\": 1024,
        \"LEARNING_RATE\": 0.00005,
        \"LORA_RANK\": 8,
        \"LORA_DROPOUT\": 0.1,
        \"LORA_TARGET\": \"all\",
        \"QUANTIZATION_BIT\": -1,
        \"GRADIENT_CHECKPOINTING\": false,
        \"GRADIENT_ACCUMULATION_STEPS\": 4,
        \"SAVE_STEPS\": 100,
        \"SAVE_TOTAL_LIMIT\": 2,
        \"LOGGING_STEP\": 10,
        \"FLASH_ATTN\": false,
        \"SHIFT_ATTN\": false,
        \"DO_EVAL\": false,
        \"EVALUATION_STRATEGY\": \"\",
        \"EVAL_STEPS\": 10,
        \"EPOCH\": 3,
        \"BATCH_SIZE\": 1,
        \"EVAL_SIZE\": 0.01,
        \"REWARD_MODEL\": \"/.reward\",
        \"PORT\": \"\"
      }",
      "batch_size": 4,
      "billing_mode": "resourcePool",

```

```

    "dataset_id": null,
    "dataset_ids": "[115]",
    "default_system": null,
    "epoch": 3,
    "exclusive": "",
    "expand": "",
    "framework": "deepspeed",
    "framework_config": "{\n
        \"train_batch_size\": \"auto\", \n
        \"train_micro_batch_size_per_gpu\": \"auto\", \n
        \"gradient_accumulation_steps\": \"auto\", \n
        \"gradient_clipping\": \"auto\", \n
        \"zero_allow_untested_optimizer\": true, \n
        \"fp16\": {\n
            \"enabled\": \"auto\", \n
            \"loss_scale\": 0, \n
            \"initial_scale_power\": 16, \n
            \"loss_scale_window\": 1000, \n
            \"hysteresis\": 2, \n
            \"min_loss_scale\": 1\n
        }, \n
        \"zero_optimization\": {\n
            \"stage\": 2, \n
            \"allgather_partitions\": true, \n
            \"allgather_bucket_size\": 5e8, \n
            \"reduce_scatter\": true, \n
            \"reduce_bucket_size\": 5e8, \n
            \"overlap_comm\": true, \n
            \"contiguous_gradients\": true\n
        }\n
    }",
    "framework_template": "stage2",
    "ft_type": "lora",
    "id": 42,
    "label": "test-002",
    "level": "L0",
    "maas_task_parent_id": null,
    "model_path": "",
    "model_type": "llama2_zh",
    "name": "test-002",
    "node_selector": "train=true,gpu=true",
    "output_dir": "",
    "resource_cpu": "50",
    "resource_gpu": "2(H800)",
    "resource_id": "AIDP-f5adef644a45ca8ab4a56c0f0f63ab",
    "resource_memory": "200G",
    "retry": 0,
    "spec_type": "compute",
    "storage_type": 0,
    "workers": 1
},
"status": 0
}
}

```

## 7.3 大模型预训练、调优(SFT、DPO、RLHF)编辑接口

### 7.3.1 简要描述

大模型预训练、调优(SFT、DPO、RLHF)任务编辑。

### 7.3.2 接口约束

无。

### 7.3.3 URL 格式

大模型任务类型	URL 格式
预训练	https://{endpoint}/api/aidp/alpha/maas_task_modelview/pt/api/{taskId}?requestInfo={regionId}
调优SFT	https://{endpoint}/api/aidp/alpha/maas_task_modelviewsft/api/{taskId}?requestInfo={regionId}
调优DPO	https://{endpoint}/api/aidp/alpha/maas_task_modelview/dpo/api/{taskId}?requestInfo={regionId}
调优 RLHF—奖励模型训练	https://{endpoint}/api/aidp/alpha/maas_task_modelview/rw/api/{taskId}?requestInfo={regionId}
调优 RLHF—强化学习训练	https://{endpoint}/api/aidp/alpha/maas_task_modelview/ppo/api/{taskId}?requestInfo={regionId}

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
{regionId}	必须	string	region可用区
{taskId}	必须	int	训练任务ID

### 7.3.4 请求方式

PUT。

### 7.3.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type	必须	-	application/json	http内容类型, 根据实际情况传参

请求 body 参数:

参数	是否必选	参数类型	示例	描述
jobname	是	string	hsk-0711	作业名称
name	是	string	test-002	任务名称
label	是	string	test-002	任务标签
storage_type	是	int	0	存储类型, 0 表示使用默认存储类型
billing_mode	是	string	resourcePool	计费模式
bill_gui	否	string	defaultBill	计费界面, 留空使用默认
spec_type	是	string	compute	规格类型, 如计算型
node_selector	是	string	train=true	节点选择器, 用于指定任务运行的节点
workers	是	int	1	工作节点数量
resource_id	是	string	AIDP-f5adef644a45ca8ab4a56c0f0f63ab	资源ID
framework	是	string	deepspeed	框架类型, 如 deepspeed
ft_type	是	string	lora	功能类型, 如 lora
level	是	string	L0	等级
advanced_mode	是	string	close	高级模式, 如关闭
framework_template	是	string	stage2	框架模板
aihub_id	是	int	41	AIHub ID
maas_task_parent_id	否	null	null	父任务ID, 留空表示无父任务
model_path	否	null	null	模型路径, 留空表示使用默认
model_type	否	null	null	模型类型, 留空表示使用默认
projectId	是	int	45	项目ID
dataset_ids	是	string	[115]	数据集ID列表
maas	是	int	44	MAAS ID
args	是	string	详细的参数JSON字符串	运行参数
framework_config	是	string	详细的框架配置JSON字符串	框架配置

## 7.3.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	返回的数据对象

### data 对象结构

参数	参数类型	描述
code	string	状态码
message	string	状态消息
result	object	任务详细信息
status	int	状态值

### result 对象结构:

参数	参数类型	描述
advanced_mode	string	任务是否启用高级模式, 例如 "close" 表示关闭高级模式
aihub_id	int	与AIHub相关联的ID
args	string	任务运行参数的JSON字符串
batch_size	int	批处理大小
billing_mode	string	计费模式, 例如 "resourcePool"
dataset_id	null	数据集ID, 返回中可能为null
dataset_ids	string	数据集ID列表, JSON格式的字符串
default_system	null	默认系统配置, 返回中可能为null
epoch	int	训练周期数
exclusive	string	独占模式, 返回中可能为空字符串
expand	string	扩展配置, 返回中可能为空字符串
framework	string	使用的框架, 例如 "deepspeed"
framework_config	string	框架配置的JSON字符串
framework_template	string	框架模板名称, 例如 "stage2"
ft_type	string	功能类型, 例如 "lora"
id	int	任务的唯一标识符
label	string	任务标签

参数	参数类型	描述
level	string	任务等级，例如 "L0"
maas_task_parent_id	null	父任务ID，返回中可能为null
model_path	string	模型路径，返回中可能为空字符串
model_type	string	模型类型，例如 "llama2_zh"
name	string	任务名称
node_selector	string	节点选择器
output_dir	string	输出目录，返回中可能为空字符串
resource_cpu	string	CPU资源规格，例如 "50"
resource_gpu	string	GPU资源规格，例如 "2(H800)"
resource_id	string	资源ID
resource_memory	string	内存资源规格，例如 "200G"
retry	int	重试次数
spec_type	string	规格类型，例如 "compute"
storage_type	int	存储类型
workers	int	工作节点数量

### 7.3.7 示例

请求示例：

```
PUT 'https://{endpoint}/api/aidp/alpha/maas_task_modelview/pt/api/42?requestInfo=region1'
--header 'Content-Type: application/json'
--data-raw '{
  "jobname": "hsk-0711",
  "name": "test-002",
  "label": "test-002",
  "storage_type": 0,
  "billing_mode": "resourcePool",
  "bill_gui": "defaultBill",
  "spec_type": "compute",
  "node_selector": "train=true",
  "workers": 1,
  "resource_id": "AIDP-f5adef644a45ca8ab4a56c0f0f63ab",
  "framework": "deepspeed",
  "ft_type": "lora",
  "level": "L0",
  "advanced_mode": "close",
  "framework_template": "stage2",
  "aihuh_id": 41,
  "maas_task_parent_id": null,
  "model_path": null,
  "model_type": null,
  "projectId": 45,
  "dataset_ids": "[115]",
  "maas": "44",
  "args": {
    "MAX_LENGTH": 1024,
    "LEARNING_RATE": 0.00005,
    "LORA_RANK": 8,
    "LORA_DROPOUT": 0.1,
    "LORA_TARGET": "all",
    "QUANTIZATION_BIT": -1,
    "GRADIENT_CHECKPOINTING": false,
    "GRADIENT_ACCUMULATION_STEPS": 4,
    "SAVE_STEPS": 100,
    "SAVE_TOTAL_LIMIT": 2,
    "LOGGING_STEP": 10,
    "FLASH_ATTN": false,
    "SHIFT_ATTN": false,
    "DO_EVAL": false,
    "EVALUATION_STRATEGY": "\",
    "EVAL_STEPS": 10,
    "EPOCH": 3,
    "BATCH_SIZE": 1,
    "EVAL_SIZE": 0.01,
    "REWARD_MODEL": "./reward",
    "PORT": "\n\n"},
  "framework_config": {
    "train_batch_size": "auto",
    "train_micro_batch_size_per_gpu":
```

```

"auto",\n"gradient_accumulation_steps":\n"auto",\n"gradient_clipping":\n"auto",\n"zero_allow_
untested_optimizer":true,\n"fp16":{\n"enabled":\n"auto",\n"loss_scale":\n0,\n
"initial_scale_power":\n16,\n"loss_scale_window":\n
1000,\n"hysteresis":2,\n"min_loss_scale":1\n},\n"zero_optimization":{\n"stage":2,\n"allgathe
r_partitions":true,\n"allgather_bucket_size":5e8,\n"reduce_scatter":true,\n"reduce_bucket_siz
e":5e8,\n"overlap_commm":true,\n"contiguous_gradients": true\n}\n}'

```

正常返回示例:

HTTP/1.1 200 OK

Content-Type:application/json

```

{
  "requestId": "539ff1f2a578478b996d5b4471e23d7d",
  "code": "Success",
  "message": "成功",
  "data": {
    "code": "",
    "message": "success",
    "result": {
      "advanced_mode": "close",
      "aihub_id": 41,
      "args": "{\n  \"MAX_LENGTH\": 1024,\n  \"LEARNING_RATE\": 0.00005,\n
      \"LORA_RANK\": 8,\n  \"LORA_DROPOUT\": 0.1,\n  \"LORA_TARGET\": \"all\",\n
      \"QUANTIZATION_BIT\": -1,\n  \"GRADIENT_CHECKPOINTING\": false,\n
      \"GRADIENT_ACCUMULATION_STEPS\": 4,\n  \"SAVE_STEPS\": 100,\n
      \"SAVE_TOTAL_LIMIT\": 2,\n  \"LOGGING_STEP\": 10,\n  \"FLASH_ATTN\": false,\n
      \"SHIFT_ATTN\": false,\n  \"DO_EVAL\": false,\n  \"EVALUATION_STRATEGY\": \"\",
      \"EVAL_STEPS\": 10,\n  \"EPOCH\": 3,\n  \"BATCH_SIZE\": 1,\n  \"EVAL_SIZE\": 0.01,\n
      \"REWARD_MODEL\": \"/.reward\",\n  \"PORT\": \"\"\n}",
      "batch_size": 4,
      "billing_mode": "resourcePool",
      "dataset_id": null,
      "dataset_ids": "[115]",
      "default_system": null,
      "epoch": 3,
      "exclusive": "",
      "expand": "",
      "framework": "deepspeed",

```



```

    "framework_config": "{\n
        \"train_batch_size\": \"auto\", \n
        \"train_micro_batch_size_per_gpu\": \"auto\", \n
        \"gradient_accumulation_steps\": \"auto\", \n
        \"gradient_clipping\": \"auto\", \n
        \"zero_allow_untested_optimizer\": true, \n
        \"fp16\": {\n
            \"enabled\": \"auto\", \n
            \"loss_scale\": 0, \n
            \"initial_scale_power\": 16, \n
            \"loss_scale_window\": 1000, \n
            \"hysteresis\": 2, \n
            \"min_loss_scale\": 1\n
        }, \n
        \"zero_optimization\": {\n
            \"stage\": 2, \n
            \"allgather_partitions\": true, \n
            \"allgather_bucket_size\": 5e8, \n
            \"reduce_scatter\": true, \n
            \"reduce_bucket_size\": 5e8, \n
            \"overlap_comm\": true, \n
            \"contiguous_gradients\": true\n
        } \n
    },\n
    "framework_template": "stage2",\n
    "ft_type": "lora",\n
    "id": 42,\n
    "label": "test-002",\n
    "level": "L0",\n
    "maas_task_parent_id": null,\n
    "model_path": "",\n
    "model_type": "llama2_zh",\n
    "name": "test-002",\n
    "node_selector": "train=true,gpu=true",\n
    "output_dir": "",\n
    "resource_cpu": "50",\n
    "resource_gpu": "2(H800)",\n
    "resource_id": "AIDP-f5adeffe644a45ca8ab4a56c0f0f63ab",\n
    "resource_memory": "200G",\n
    "retry": 0,\n
    "spec_type": "compute",\n
    "storage_type": 0,\n
    "workers": 1\n
},\n
"status": 0\n
}\n
}

```

## 7.4 大模型预训练、调优(SFT、DPO、RLHF)详情接口

### 7.4.1 简要描述

大模型预训练、调优(SFT、DPO、RLHF)详情查询。

## 7.4.2 接口约束

无。

## 7.4.3 URL 格式

大模型任务类型	URL 格式
预训练	https://{endpoint}/api/aidp/alpha/maas_task_modelview/pt/api/{taskId}?requestInfo={regionId}
调优SFT	https://{endpoint}/api/aidp/alphamaas_task_modelview/sft/api/{taskId}?requestInfo={regionId}
调优DPO	https://{endpoint}/api/aidp/alpha/maas_task_modelview/dpo/api/{taskId}?requestInfo={regionId}
调优 RLHF—奖励模型训练	https://{endpoint}/api/aidp/alpha/maas_task_modelview/rw/api/{taskId}?requestInfo={regionId}
调优 RLHF—强化学习训练	https://{endpoint}/api/aidp/alpha/maas_task_modelview/ppo/api/{taskId}?requestInfo={regionId}

URL 参数说明：

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
{taskId}	必须	int	任务id
{regionId}	必须	string	region可用区

## 7.4.4 请求方式

GET。

## 7.4.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type	必须	-	application/json	http内容类型，根据实际情况传参

## 7.4.6 返回参数

参数	参数类型	描述
message	string	状态消息

参数	参数类型	描述
result	object	任务详细信息
status	int	状态值

### result 对象结构

参数	参数类型	描述
advanced_mode	string	模型的高级模式
aihub.dataset	object	数据集描述, 可为空
aihub.describe	string	模型描述,
aihub.doc	string	模型文档链接, 可为空
aihub.expand	object	扩展信息
aihub.field	string	字段描述
aihub.hot	int	热度值
aihub.id	int	模型ID
aihub.images	string	图片链接, 值可为空
aihub.inference	object	推理信息, 值可为空
aihub.inference_service_id	null	推理服务ID, 值可为空
aihub.job_template	object	作业模板信息, 值可为空
aihub.label	string	标签
aihub.model_class	string	模型类别
aihub.model_class_type	string	模型类别类型, 值可为 "chat"
aihub.model_file_size	string	模型文件大小
aihub.model_max_inference_length	int	最大推理长度, 值可为 8192
aihub.model_parameters	object	模型参数配置, 包含资源配置
aihub.model_path	string	模型路径
aihub.model_source	string	模型来源, 值可为 "aihub"
aihub.model_type	string	模型类型
aihub.name	string	名称
aihub.notebook	object	notebook信息, 值可为空
aihub.parameter_quantity	string	参数数量,
aihub.pic	string	图片文件名
aihub.pipeline	string	流水线信息
aihub.pre_train_model	object	预训练模型信息

参数	参数类型	描述
aihub.price	int	价格
aihub.quantization_bit	string	量化位宽, 值可为 "fp16"
aihub.scenes	string	应用场景
aihub.service	object	服务信息
aihub.source	string	来源信息
aihub.status	string	状态
aihub.type	string	类型, 为 "dataset,notebook,train,inference"
aihub.uuid	string	唯一标识符
aihub.version	string	版本信息
args	object	参数配置, 包含训练和推理时的参数设置
batch_size	int	批处理大小,
billing_mode	string	计费模式
changed_on	string	最后修改时间
created_on	string	创建时间
creator_first_name	string	创建者名字
dataset	string	数据集名称
dataset_id	null	数据集ID
dataset_ids	string	数据集IDs列表
default_system	null	默认系统配置
epoch	float	训练周期数
exclusive	string	资源独占配置
expand	string	扩展信息
framework	string	框架类型, 为 "deepspeed"
framework_config	object	框架配置, 包含训练和优化参数
framework_template	string	框架模板, 可为 "stage2"
ft_type	string	微调类型
id	int	任务ID
label	string	标签
last_run_time	string	最后运行时间
level	string	等级
lock_status	int	锁定状态
maas	object	MAAS相关配置, 包含多个字段

参数	参数类型	描述
maas_task_parent	null	MAAS任务父级
maas_task_parent_id	null	MAAS任务父级ID
model_path	string	模型路径
model_type	string	模型类型，为 "chatglm3"
name	string	名称
node_selector	string	节点选择器
output_dir	string	输出目录
resource_cpu	string	CPU资源
resource_gpu	string	GPU资源
resource_id	string	资源ID
resource_memory	string	内存资源
retry	int	重试次数
spec_type	string	规格类型
status	string	状态
storage_type	int	存储类型
workers	int	工作进程数

## 7.4.7 示例

请求示例：

```
GET'https://{endpoint}/api/aidp/alpha/maas_task_modelview/pt/api/8?requestInfo=region1'
--header 'Content-Type: application/json'
```

正常返回示例：

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "message": "success",
  "result": {
    "advanced_mode": "close",
    "aihub": {
      "dataset": "{}",
      "describe": "chatglm3-6b",
```

```

"doc": "https://www.modelscope.cn/models/ZhipuAI/chatglm3-6b/files",
"expand": "{}",
"field": "\u5927\u6a21\u578b",
"hot": 24427,
"id": 42,
"images": "",
"inference": "{}",
"inference_service_id": null,
"job_template": "{}",
"label": "chatglm3-6b",
"model_class": "LLM",
"model_class_type": "chat",
"model_file_size": "13GB",
"model_max_inference_length": "8192",
"model_parameters": "{n  \"resource_gpu\": \"1\",n  \"resource_memory\": \"40G\",n
\"resource_cpu\": \"4\"n}",
"model_path": "/mnt/workspace/models/public-models/llm/chatglm3-6b",
"model_source": "aihub",
"model_type": "chatglm3",
"name": "chatglm3-6b",
"notebook": "{}",
"parameter_quantity": "6B",
"pic": "example.jpeg",
"pipeline": "",
"pre_train_model": "{}",
"price": 0,
"quantization_bit": "fp16",
"scenes": "\u6587\u672c\u751f\u6210",
"service": "{}",
"source": "",
"status": "online",
"type": "dataset,notebook,train,inference",
"uuid": "chatglm3-6b",
"version": ""
},
"aihub_id": 42,
"args": "{n  \"MAX_LENGTH\": 1024,n  \"LEARNING_RATE\": 0.00005,n  \"LORA_RANK\":
8,n  \"LORA_DROPOUT\": 0.1,n  \"LORA_TARGET\": \"all\",n  \"QUANTIZATION_BIT\": -1,n

```

```

\ "GRADIENT_CHECKPOINTING": false,\n  \ "GRADIENT_ACCUMULATION_STEPS": 4,\n
\ "SAVE_STEPS": 100,\n  \ "SAVE_TOTAL_LIMIT": 2,\n  \ "LOGGING_STEP": 10,\n
\ "FLASH_ATTN": false,\n  \ "SHIFT_ATTN": false,\n  \ "DO_EVAL": false,\n
\ "EVALUATION_STRATEGY": "",\n  \ "EVAL_STEPS": 10,\n  \ "EPOCH": 3,\n  \ "BATCH_SIZE":
1,\n  \ "EVAL_SIZE": 0.01,\n  \ "REWARD_MODEL": "/reward",\n  \ "PORT": "\n"},
  "batch_size": 4,
  "billing_mode": "resourcePool",
  "changed_on": "2024-07-11T20:59:14.957932",
  "created_on": "2024-06-19T14:42:42",
  "creator_first_name": "guodx1",
  "dataset": "\u5c0f\u8bf4",
  "dataset_id": null,
  "dataset_ids": "[103]",
  "default_system": null,
  "epoch": 3.0,
  "exclusive": "",
  "expand": "",
  "framework": "deepspeed",
  "framework_config": "{\n
    \ "train_batch_size": \ "auto",\n
\ "train_micro_batch_size_per_gpu": \ "auto",\n  \ "gradient_accumulation_steps": \ "auto",\n
\ "gradient_clipping": \ "auto",\n  \ "zero_allow_untested_optimizer": true,\n  \ "fp16":
{\n    \ "enabled": \ "auto",\n    \ "loss_scale": 0,\n    \ "initial_scale_power":
16,\n    \ "loss_scale_window": 1000,\n    \ "hysteresis": 2,\n    \ "min_loss_scale":
1\n  },\n  \ "zero_optimization": {\n    \ "stage": 2,\n    \ "allgather_partitions":
true,\n    \ "allgather_bucket_size": 5e8,\n    \ "reduce_scatter": true,\n
\ "reduce_bucket_size": 5e8,\n    \ "overlap_comm": true,\n    \ "contiguous_gradients":
true\n  }\n  }",
  "framework_template": "stage2",
  "ft_type": "full",
  "id": 8,
  "label": "test-111",
  "last_run_time": "\u672a\u77e5",
  "level": "L0",
  "lock_status": 0,
  "maas": {
    "alert_status": "Pending,Running,Succeeded,Failed,Terminated",
    "alert_user": "",
    "changed_on": "2024-06-11T18:34:59",
    "created_on": "2024-06-11T18:34:59",

```

```
"cron_time": null,
"cronjob_start_time": "2024-06-11 18:34:58",
"dag_json": "{}",
"depends_on_past": false,
"describe": "1",
"expand": "[]",
"expired_limit": 0,
"global_env": "",
"id": 5,
"image_pull_policy": "IfNotPresent",
"max_active_runs": 3,
"name": "test-hr11",
"namespace": "pipeline",
"node_selector": "cpu=true,train=true",
"parallelism": 1,
"parameter": "{}",
"pipeline_argo_id": null,
"pipeline_file": "",
"run_id": null,
"schedule_type": "once",
"type": "pt",
"version_id": null
},
"maas_task_parent": null,
"maas_task_parent_id": null,
"model_path": "",
"model_type": "chatglm3",
"name": "test-111",
"node_selector": "train=true,gpu=true",
"output_dir": "",
"resource_cpu": "64",
"resource_gpu": "4(H800)",
"resource_id": "AIDP-a1a59b9d712249d08a61e848900ec620",
"resource_memory": "1024",
"retry": 0,
"spec_type": "compute",
"status": "",
"storage_type": 0,
```



```

    "workers": 4
  },
  "status": 0
}

```

## 7.5 大模型预训练、调优(SFT、DPO、RLHF)列表接口

### 7.5.1 简要描述

大模型预训练、调优(SFT、DPO、RLHF)任务列表查询。

### 7.5.2 接口约束

无。

### 7.5.3 URL 格式

大模型任务类型	URL 格式
预训练	https://{endpoint}/api/aidp/alpha/maas_task_modelview/pt/api?requestInfo={regionId}
调优SFT	https://{endpoint}/api/aidp/alpha/maas_task_modelview/sft/api?requestInfo={regionId}
调优DPO	https://{endpoint}/api/aidp/alpha/maas_task_modelview/dpo/api?requestInfo={regionId}
调优RLHF—奖励模型训练	https://{endpoint}/api/aidp/alpha/maas_task_modelview/rw/api?requestInfo={regionId}
调优RLHF—强化学习训练	https://{endpoint}/api/aidp/alpha/maas_task_modelview/ppo/api?requestInfo={regionId}

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
{regionId}	必须	string	region可用区

### 7.5.4 请求方式

GET。

### 7.5.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type	必须	-	application/json	http内容类型，根据实际情况传参

请求 body 参数:

参数	是否必选	参数类型	示例	描述
requestInfo	否	string	region1	请求信息标识符。
form_data	是	string	{"page":0,"page_size":10,...}	JSON格式的查询参数。

## 7.5.6 返回参数

参数	参数类型	描述
message	string	请求结果描述
result	object	包含查询结果的数组
status	int	请求状态码

result 对象结构:

参数	参数类型	描述
count	int	数据条目总数。
data	[]object	包含具体数据的数组

data 数组元素结构

参数	参数类型	描述
aihub	string	AIHub实例的名称。
aihub_id	int	AIHub实例的唯一标识符ID。
args	string	任务参数，JSON格式的字符串。
billing_mode	string	计费模式。
changed_by	string	最后修改任务的用户ID。
changed_on	string	最后修改任务的时间。
created_by	string	创建任务的用户ID。
created_on	string	创建任务的时间。
creator_first_name	string	创建任务的用户的名字。
dataset	string	数据集名称。
dataset_id	object	数据集ID（如果为null，则可能表示未指定或不适用）。

参数	参数类型	描述
dataset_ids	string	包含数据集ID的数组，JSON格式的字符串。
elapsed_time	string	任务已运行时间，格式为"00小时00分钟"。
final_status	string	任务的最终状态，例如"Failed"。
ft_type	string	微调类型，例如"full"。
id	int	任务的唯一标识符ID。
label	string	任务的标签。
last_run_time	string	任务最后一次运行的时间。
level	string	任务的级别，例如"L0"。
lock_status	int	任务的锁定状态。
maas	string	MaaS实例的名称。
maas_task_parent	object	父MaaS任务（如果为null，则表示没有父任务）。
maas_task_parent_id	object	父MaaS任务的ID（如果为null，则表示没有父任务）。
model_path	string	模型路径。
model_type	string	模型类型。
name	string	任务的名称。
node_selector	string	节点选择器，用于指定任务运行的节点。
percentage	string	任务完成的百分比，例如"--"表示未完成。
resource_cpu	string	任务使用的CPU资源，例如"32"。
resource_gpu	string	任务使用的GPU资源，例如"2(T4)"。
resource_id	string	任务使用的资源ID。
resource_memory	string	任务使用的内存资源，例如"256G"。
status	string	任务的当前状态，例如"Failed"。
storage_type	int	存储类型。

## 7.5.7 示例

请求示例：

GET

```
'https://cecstackzsarm.user.cecloud.com/api/aidp/alpha/maas_task/pt/api/?requestInfo=region1&form_data=%7B%22page%22:0,%22page_size%22:10,%22str_related%22:1,%22filters%22:%5B%7B%22col%22:%22maas%22,%22opr%22:%22rel_o_m%22,%22value%22:%2244%22%7D%5D'
```

正常返回示例：

HTTP/1.1 200 OK

Content-Type:application/json

```
{
  "message": "success",
  "result": {
    "count": 1,
    "data": [
      {
        "aihub": "chatglm3-6b",
        "aihub_id": 42,
        "args": "{\n  \"MAX_LENGTH\": 1024,\n  \"LEARNING_RATE\": 0.00005,\n  \"LORA_RANK\": 8,\n  \"LORA_DROPOUT\": 0.1,\n  \"LORA_TARGET\": \"all\",\n  \"QUANTIZATION_BIT\": -1,\n  \"GRADIENT_CHECKPOINTING\": false,\n  \"GRADIENT_ACCUMULATION_STEPS\": 4,\n  \"SAVE_STEPS\": 100,\n  \"SAVE_TOTAL_LIMIT\": 2,\n  \"LOGGING_STEP\": 10,\n  \"FLASH_ATTN\": false,\n  \"SHIFT_ATTN\": false,\n  \"DO_EVAL\": false,\n  \"EVALUATION_STRATEGY\": \"\",\n  \"EVAL_STEPS\": 10,\n  \"EPOCH\": 3,\n  \"BATCH_SIZE\": 1,\n  \"EVAL_SIZE\": 0.01,\n  \"REWARD_MODEL\": \"./reward\",\n  \"PORT\": \"\\n\"}",
        "billing_mode": "resourcePool",
        "changed_by": "1841781662849977755",
        "changed_on": "2024-07-11T19:44:21",
        "created_by": "1841781662849977755",
        "created_on": "2024-07-11T19:07:02",
        "creator_first_name": "guodx1",
        "dataset": "pt-train",
        "dataset_id": null,
        "dataset_ids": "[115]",
        "elapsed_time": "00\u5c0f\u65f6\u5206\u949f",
        "final_status": "Failed",
        "ft_type": "full",
        "id": 40,
        "label": "xxx",
        "last_run_time": "2024-07-11 19:44:21",
        "level": "L0",
        "lock_status": 0,
        "maas": "hsk-0711",
        "maas_task_parent": null,
        "maas_task_parent_id": null,
        "model_path": ""
      }
    ]
  }
}
```

```

    "model_type": "chatglm3",
    "name": "hsk-pt1",
    "node_selector": "train=true,gpu=true",
    "percentage": "--",
    "resource_cpu": "32",
    "resource_gpu": "2(T4)",
    "resource_id": "AIDP-e5a35b75955c4b179a74d5911cf701dd",
    "resource_memory": "256G",
    "status": "Failed",
    "storage_type": 0
  }
]
},
"status": 0
}

```

## 7.6 大模型预训练、调优(SFT、DPO、RLHF)运行接口

### 7.6.1 简要描述

大模型预训练、调优(SFT、DPO、RLHF)任务运行。

### 7.6.2 接口约束

无。

### 7.6.3 URL 格式

大模型任务类型	URL 格式
预训练	https://{endpoint}/api/aidp/alpha/maas_task_modelview/pt/api/maas_run/{taskId}
调优SFT	https://{endpoint}/api/aidp/alpha/maas_task_modelview/sft/api/maas_run/{taskId}
调优DPO	https://{endpoint}/api/aidp/alpha/maas_task_modelview/dpo/api/maas_run/{taskId}
调优RLHF—奖励模型训练	https://{endpoint}/api/aidp/alpha/maas_task_modelview/rw/api/maas_run/{taskId}
调优RLHF—强化学习训练	https://{endpoint}/api/aidp/alpha/maas_task_modelview/ppo/api/maas_run/{taskId}

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
{taskId}	必须	int	任务id

## 7.6.4 请求方式

GET。

## 7.6.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type	是	-	application/json	http内容类型，根据实际情况传参

## 7.6.6 返回参数

参数	参数类型	描述
message	string	信息描述
result	string	任务运行的执行信息
status	int	状态值

## 7.6.7 示例

请求示例：

```
GET 'https://{endpoint}/api/aidp/alpha/maas_task_modelview/pt/api/maas_run/40'
--header 'Content-Type: application/json'
```

正常返回示例：

```
HTTP/1.1 200 OK
Content-Type:application/json
{"message":"success","result":{},"status":0}
```

## 7.7 大模型预训练、调优(SFT、DPO、RLHF)停止接口

### 7.7.1 简要描述

大模型预训练、调优(SFT、DPO、RLHF)任务停止。

## 7.7.2 接口约束

无。

## 7.7.3 URL 格式

大模型任务类型	URL 格式
预训练	https://{endpoint}/api/aidp/alpha/maas_task_modelview/pt/api/stop/{taskId}?requestInfo={regionId}
调优 SFT	https://{endpoint}/api/aidp/alpha/maas_task_modelview/sft/api/stop/{taskId}?requestInfo={regionId}
调优 DPO	https://{endpoint}/api/aidp/alpha/maas_task_modelview/dpo/api/stop/{taskId}?requestInfo={regionId}
调优 RLHF —奖励模型训练	https://{endpoint}/api/aidp/alpha/maas_task_modelview/rw/api/stop/{taskId}?requestInfo={regionId}
调优 RLHF —强化学习训练	https://{endpoint}/api/aidp/alpha/maas_task_modelview/ppo/api/stop/{taskId}?requestInfo={regionId}

URL 参数说明：

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
{taskId}	必须	int	任务id
{regionId}	必须	string	region可用区

## 7.7.4 请求方式

GET。

## 7.7.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type	是	-	application/json	http内容类型，根据实际情况传参

## 7.7.6 返回参数

参数	参数类型	描述
message	string	信息描述
result	string	任务运行的执行信息
status	int	状态值

## 7.7.7 示例

请求示例：

```
GET 'https://{endpoint}/api/aidp/alpha/maas_task_modelview/pt/api/stop/40'  
--header 'Content-Type: application/json'
```

正常返回示例：

```
HTTP/1.1 200 OK  
Content-Type:application/json  
{  
  "message": "success",  
  "result": {},  
  "status": 0  
}
```

## 7.8 大模型预训练、调优(SFT、DPO、RLHF)删除接口

### 7.8.1 简要描述

大模型预训练、调优(SFT、DPO、RLHF)任务删除。

### 7.8.2 接口约束

无。

### 7.8.3 URL 格式

`https://{endpoint}/api/aidp/v1/deleteInstance?requestInfo={regionId}&id={taskId}&type={taskType}`

URL 参数说明：

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
{regionId}	必须	string	region可用区
{taskId}	必须	int	任务id
{taskType}	必须	int	任务类型，大模型任务为5



## 7.8.4 请求方式

DELETE。

## 7.8.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type	必须	-	application/json	http内容类型，根据实际情况传参

## 7.8.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	返回的数据对象

data 对象结构：

参数	参数类型	描述
code	string	状态码
message	string	状态消息
result	object	任务详细信息
status	int	状态值

result 对象结构：

参数	参数类型	描述
advanced_mode	string	任务是否启用高级模式，例如 "close" 表示关闭高级模式
aihub_id	int	与AIHub相关联的ID
args	string	任务运行参数的JSON字符串
batch_size	int	批处理大小
billing_mode	string	计费模式，例如 "resourcePool"
dataset_id	null	数据集ID，返回中可能为null
dataset_ids	string	数据集ID列表，JSON格式的字符串
default_system	null	默认系统配置，返回中可能为null

参数	参数类型	描述
epoch	int	训练周期数
exclusive	string	独占模式，返回中可能为空字符串
expand	string	扩展配置，返回中可能为空字符串
framework	string	使用的框架，例如 "deepspeed"
framework_config	string	框架配置的JSON字符串
framework_template	string	框架模板名称，例如 "stage2"
ft_type	string	功能类型，例如 "lora"
id	int	任务的唯一标识符
label	string	任务标签
level	string	任务等级，例如 "L0"
maas_task_parent_id	null	父任务ID，返回中可能为null
model_path	string	模型路径，返回中可能为空字符串
model_type	string	模型类型，例如 "llama2_zh"
name	string	任务名称
node_selector	string	节点选择器
output_dir	string	输出目录，返回中可能为空字符串
resource_cpu	string	CPU资源规格，例如 "50"
resource_gpu	string	GPU资源规格，例如 "2(H800)"
resource_id	string	资源ID
resource_memory	string	内存资源规格，例如 "200G"
retry	int	重试次数
spec_type	string	规格类型，例如 "compute"
storage_type	int	存储类型
workers	int	工作节点数量

## 7.8.7 示例

请求示例：

```
DELETE 'https://{endpoint}/api/aidp/v1/deleteInstance?requestInfo=region1&id=40&type=5' \
--header 'Content-Type: application/json'
,
```

正常返回示例：

```
HTTP/1.1 200 OK
```

Content-Type:application/json

```
{
  "requestId": "2e2ca52b45b04fba93c722eed6cc897",
  "code": "Success",
  "message": "成功",
  "data": {
    "code": "",
    "message": "success",
    "result": {
      "advanced_mode": "close",
      "aihub_id": 42,
      "alert_status": "Pending,Running,Succeeded,Failed,Terminated",
      "alert_user": "",
      "args": "{\n  \"MAX_LENGTH\": 1024,\n  \"LEARNING_RATE\": 0.00005,\n  \"LORA_RANK\": 8,\n  \"LORA_DROPOUT\": 0.1,\n  \"LORA_TARGET\": \"all\",\n  \"QUANTIZATION_BIT\": -1,\n  \"GRADIENT_CHECKPOINTING\": false,\n  \"GRADIENT_ACCUMULATION_STEPS\": 4,\n  \"SAVE_STEPS\": 100,\n  \"SAVE_TOTAL_LIMIT\": 2,\n  \"LOGGING_STEP\": 10,\n  \"FLASH_ATTN\": false,\n  \"SHIFT_ATTN\": false,\n  \"DO_EVAL\": false,\n  \"EVALUATION_STRATEGY\": \"\",\n  \"EVAL_STEPS\": 10,\n  \"EPOCH\": 3,\n  \"BATCH_SIZE\": 1,\n  \"EVAL_SIZE\": 0.01,\n  \"REWARD_MODEL\": \"/.reward\",\n  \"PORT\": \"\"\n}",
      "batch_size": 4,
      "billing_mode": "resourcePool",
      "changed_by_fk": 14,
      "changed_on": "2024-07-11T21:03:47",
      "created_by_fk": 14,
      "created_on": "2024-07-11T19:07:02",
      "dataset_id": null,
      "dataset_ids": "[115]",
      "default_system": null,
      "epoch": 3,
      "exclusive": "",
      "expand": "",
      "framework": "deepspeed",
      "framework_config": "{\n  \"train_batch_size\": \"auto\",\n  \"train_micro_batch_size_per_gpu\": \"auto\",\n  \"gradient_accumulation_steps\": \"auto\",\n  \"gradient_clipping\": \"auto\",\n  \"zero_allow_untested_optimizer\": true,\n  \"fp16\": {\n    \"enabled\": \"auto\"\n  },\n  \"zero_optimization\": {\n    \"stage\": 2,\n    \"allgather_partitions\": true,\n    \"allgather_bucket_size\": 5e8,\n
```

```

\"reduce_scatter\": true,\n          \"reduce_bucket_size\": 5e8,\ntrue,\n          \"contiguous_gradients\": true\n        }\n      }\",
    \"framework_template\": \"自定义配置参数\",
    \"ft_type\": \"lora\",
    \"id\": 40,
    \"inference_service_id\": null,
    \"is_deleted\": \"true\",
    \"job_template_id\": null,
    \"label\": \"xxx\",
    \"level\": \"L0\",
    \"lock_status\": 0,
    \"maas_id\": 44,
    \"maas_task_parent_id\": null,
    \"maas_task_rw_id\": null,
    \"model\": \"\",
    \"model_path\": \"\",
    \"model_type\": \"chatglm3\",
    \"monitoring\": \"{}\",
    \"name\": \"hsk-pt1\",
    \"node_selector\": \"train=true,gpu=true\",
    \"output_dir\": \"/mnt/workspace/maas/pt/44/40\",
    \"outputs\": \"{}\",
    \"overwrite_entrypoint\": false,
    \"rdma\": false,
    \"resource_cpu\": \"32\",
    \"resource_gpu\": \"3(T4)\",
    \"resource_id\": \"AIDP-f11ee4681a6143e494eab95555c4edba\",
    \"resource_memory\": \"256G\",
    \"resource_rdma\": \"0\",
    \"retry\": 0,
    \"run_id\": \"run-44-40\",
    \"skip\": false,
    \"spec_type\": \"compute\",
    \"storage_type\": 0,
    \"task_argo_id\": \"hsk-pt1-c4e7\",
    \"timeout\": 0,
    \"version_id\": \"\",
    \"volume_mount\": \"kubeflow-user-workspace(pvc):/mnt\",

```

```

        "workers": 1
    },
    "status": 0
}
}

```

## 7.9 大模型评估任务创建接口

### 7.9.1 简要描述

大模型评估任务创建。

### 7.9.2 接口约束

无。

### 7.9.3 URL 格式

https://{endpoint}/api/aidp/alpha/maas\_evaluation\_modelview/api/evaluation/save?requestInfo=region1

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
{regionId}	必须	string	region可用区

### 7.9.4 请求方式

POST。

### 7.9.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type	是	-	application/json	http内容类型, 根据实际情况传参

请求 body 参数:

参数	是否必选	参数类型	示例	描述
name	是	string	"test-0001"	评估模型的名称
description	否	string	""	评估模型的描述

参数	是否必选	参数类型	示例	描述
storage_type	是	string	"0"	存储类型，具体含义需进一步明确
spec_type	是	string	"compute"	规格类型，如计算型或内存型
type	是	int	2	模型类型，具体取值需进一步明确
datasetId	是	int	115	数据集的唯一标识符
datasetName	是	string	"我的数据集>pt-train"	数据集的名称
aihub	是	int	41	AIHub ID，用于标识AIHub资源
maas_task_parent	否	string	""	父任务ID，用于任务关联
model_type	否	string	""	模型类型，具体含义需进一步明确
model_path	否	string	""	模型路径，用于指定模型文件位置
evalStandardConf	是	object	见下文	评估标准配置，包含评估模式等信息
projectId	是	int	45	项目ID，用于标识项目
specCode	是	string	"calculate2H80050c"	规格代码，用于指定计算资源配置
specCPU	是	string	"50"	CPU规格，表示核心数
specGpu	是	string	"2(H800)"	GPU规格，表示GPU型号和数量
specMemory	是	string	"200"	内存规格，单位为GB
task_specCPU	是	string	"2"	任务CPU规格，表示任务使用的核心数
task_specGpu	是	string	"1"	任务GPU规格，表示任务使用的GPU数量
task_specMemory	是	string	"8"	任务内存规格，单位为GB
task_specCode	是	string	"standardPlus"	任务规格代码，用于指定任务资源配置
resource_id	是	string	"AIDP-f023f1fa02384c15b8b4bc816f14a81d"	资源ID，用于标识计算资源
service_resource_id	是	string	"AIDP-f5adef644a45ca8ab4a56c0f0f63ab"	服务资源ID，用于标识服务资源

evalStandardConf 参数:

参数	是否必选	参数类型	示例	描述
evalMode	是	string	"manual"	评估模式
stopWordsPath	否	string	""	停用词路径
evaluationDimension	是	array	见下文	评估维度配置

evaluationDimension 参数:

参数	是否必选	参数类型	示例	描述
dimension	是	string	"满意度", "真实性"	评估维度名称
description	否	string	""	维度描述
key	是	int	1, 2	维度键值

## 7.9.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	返回的数据对象

data 对象结构:

参数	参数类型	描述
code	string	状态码
message	string	状态消息
result	object	任务详细信息
status	int	状态值

## 7.9.7 示例

请求示例:

POST

```
'https://{endpoint}/api/aidp/alpha/maas_evaluation_modelview/api/evaluation/save?requestInfo=region1'
```

```
--header 'Content-Type: application/json'
```

```
--data-raw                                '{"name":"test-0001","description":"","storage_type":"0","spec_type":"compute","type":2,"datasetId":115,"datasetName":"我的数据集>pt-train","aihub":41,"maas_task_parent":"","model_type":"","model_path":"","evalStandardConf":{"evalMode":"manual","stopWordsPath":"","evaluationDimension":[{"dimension":"满意度","description":"","key":1},{"dimension":"真实性","description":"","key":2}]},"projectId":45,"specCode":"calculate2H80050c","specCPU":"50","specGpu":"2(H800)","specMemory":"200","task_specCPU":"2","task_specGpu":"1","task_specMemory":"8","task_specCode":"standardPlus","resource_id":"AIDP-f023f1fa02384c15b8b4bc816f14a81d","service_resource_id":"AIDP-f5adef644a45ca8ab4a56c0f0f63ab"}
```

正常返回示例:

HTTP/1.1 200 OK

Content-Type:application/json

```
{
  "requestId": "3c44e8e501ca454596c49cf2a5a71901",
  "code": "Success",
  "message": "成功",
  "data": {
    "code": "",
    "message": "success",
    "result": {},
    "status": 0
  }
}
```

## 7.10 大模型评估任务编辑接口

### 7.10.1 简要描述

大模型评估任务编辑。

### 7.10.2 接口约束

无。

### 7.10.3 URL 格式

[https://{endpoint}/api/aidp/alpha/maas\\_evaluation\\_modelview/api/evaluation/save?requestInfo={regionId}](https://{endpoint}/api/aidp/alpha/maas_evaluation_modelview/api/evaluation/save?requestInfo={regionId})

URL 参数说明:



参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
{regionId}	必须	string	region可用区

## 7.10.4 请求方式

POST。

## 7.10.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type	是	-	application/json	http内容类型，根据实际情况传参

请求 body 参数:

参数	是否必选	参数类型	示例	描述
name	是	string	"test-0001"	评估模型的名称
description	否	string	""	评估模型的描述
storage_type	是	string	"0"	存储类型，具体含义需进一步明确
spec_type	是	string	"compute"	规格类型，如计算型或内存型
type	是	int	2	模型类型，具体取值需进一步明确
datasetId	是	int	115	数据集的唯一标识符
datasetName	是	string	"我的数据集>pt-train"	数据集的名称
aihub	是	int	41	AIHub ID，用于标识AIHub资源
maas_task_parent	否	string	""	父任务ID，用于任务关联
model_type	否	string	""	模型类型，具体含义需进一步明确
model_path	否	string	""	模型路径，用于指定模型文件位置
evalStandardConf	是	object	见下文	评估标准配置，包含评估模式等信息
projectId	是	int	45	项目ID，用于标识项目
specCode	是	string	"calculate2H80050c"	规格代码，用于指定计算资源配置

参数	是否必选	参数类型	示例	描述
specCPU	是	string	"50"	CPU规格，表示核心数
specGpu	是	string	"2(H800)"	GPU规格，表示GPU型号和数量
specMemory	是	string	"200"	内存规格，单位为GB
task_specCPU	是	string	"2"	任务CPU规格，表示任务使用的核心数
task_specGpu	是	string	"1"	任务GPU规格，表示任务使用的GPU数量
task_specMemory	是	string	"8"	任务内存规格，单位为GB
task_specCode	是	string	"standardPlus"	任务规格代码，用于指定任务资源配置
resource_id	是	string	"AIDP-f023f1fa02384c15b8b4bc816f14a81d"	资源ID，用于标识计算资源
service_resource_id	是	string	"AIDP-f5adeffe644a45ca8ab4a56c0f0f63ab"	服务资源ID，用于标识服务资源

evalStandardConf 参数:

参数	是否必选	参数类型	示例	描述
evalMode	是	string	"manual"	评估模式
stopWordsPath	否	string	""	停用词路径
evaluationDimension	是	array	见下文	评估维度配置

evaluationDimension 参数:

参数	是否必选	参数类型	示例	描述
dimension	是	string	"满意度", "真实性"	评估维度名称
description	否	string	""	维度描述
key	是	int	1, 2	维度键值

## 7.10.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码

参数	参数类型	描述
message	string	请求结果描述
data	object	返回的数据对象

data 对象结构:

参数	参数类型	描述
code	string	状态码
message	string	状态消息
result	object	任务详细信息
status	int	状态值

## 7.10.7 示例

请求示例:

PUT

```
'https://{endpoint}/api/aidp/alpha/maas_evaluation_modelview/api/evaluation/save?requestInfo=region1'
```

```
--header 'Content-Type: application/json'
```

```
--data-raw '{"name":"test-0001","description":"","storage_type":"0","spec_type":"compute","type":2,"datasetId":115,"datasetName":"我的数据集>pt-train","aihub":41,"maas_task_parent":"","model_type":"","model_path":"","evalStandardConf":{"evalMode":"manual","stopWordsPath":"","evaluationDimension":[{"dimension":"满意度","description":"","key":1},{"dimension":"真实性","description":"","key":2}],"projectId":45,"specCode":"calculate2H80050c","specCPU":"50","specGpu":"2(H800)","specMemory":"200","task_specCPU":"2","task_specGpu":"1","task_specMemory":"8","task_specCode":"standardPlus","resource_id":"AIDP-f023f1fa02384c15b8b4bc816f14a81d","service_resource_id":"AIDP-f5adef644a45ca8ab4a56c0f0f63ab"}'
```

正常返回示例:

```
HTTP/1.1 200 OK
```

```
Content-Type: application/json
```

```
{
  "requestId": "3c44e8e501ca454596c49cf2a5a71901",
  "code": "Success",
  "message": "成功",
  "data": {
    "code": "",
```

```
"message": "success",
"result": {},
"status": 0
}
```

## 7.11 大模型评估任务运行接口

### 7.11.1 简要描述

大模型评估任务运行。

### 7.11.2 接口约束

无。

### 7.11.3 URL 格式

https://{endpoint}/api/aidp/alpha/maas\_evaluation\_modelview/api/evaluation/start?requestInfo={regionId}

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
{regionId}	必须	string	region可用区

### 7.11.4 请求方式

POST。

### 7.11.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type	是	-	application/json	http内容类型, 根据实际情况传参

请求 body 参数:

参数	是否必选	参数类型	示例	描述
id	是	int	68	评估任务id

## 7.11.6 返回参数

参数	参数类型	描述
message	string	信息描述
result	string	任务运行的执行信息
status	int	状态值

## 7.11.7 示例

请求示例：

GET

```
'https://{endpoint}/api/aidp/alpha/maas_evaluation_modelview/api/evaluation/start?requestInfo=region1'
```

```
--header 'Content-Type: application/json'
```

```
--data-raw '{"id":68}'
```

正常返回示例：

HTTP/1.1 200 OK

Content-Type:application/json

```
{"message":"success","result":{},"status":0}
```

## 7.12 大模型评估任务停止接口

### 7.12.1 简要描述

大模型评估任务运行。

### 7.12.2 接口约束

无。

### 7.12.3 URL 格式

```
https://{endpoint}/api/aidp/alpha/maas_evaluation_modelview/api/evaluation/stop?requestInfo={regionId}
```

URL 参数说明：

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
{regionId}	必须	string	region可用区

## 7.12.4 请求方式

GET。

## 7.12.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type	是	-	application/json	http内容类型, 根据实际情况传参

请求 body 参数:

参数	是否必选	参数类型	示例	描述
id	是	int	68	评估任务id

## 7.12.6 返回参数

参数	参数类型	描述
message	string	信息描述
result	string	任务运行的执行信息
status	int	状态值

## 7.12.7 示例

请求示例:

GET

```
'https://{endpoint}/api/aidp/alpha/maas_evaluation_modelview/api/evaluation/stop?requestInfo=region1'
```

```
--header 'Content-Type: application/json'
```

```
--data-raw '{"id":68}'
```

正常返回示例:

```
HTTP/1.1 200 OK
```

```
Content-Type:application/json
```

```
{"message":"success","result":{},"status":0}
```

## 7.13 大模型评估任务删除接口

### 7.13.1 简要描述

大模型评估任务删除。

## 7.13.2 接口约束

无。

## 7.13.3 URL 格式

https://{endpoint}/api/aidp/alpha/maas\_evaluation\_modelview/api/evaluation/delete?requestInfo={regionId}

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
{regionId}	必须	string	region可用区

## 7.13.4 请求方式

GET。

## 7.13.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type	是	-	application/json	http内容类型, 根据实际情况传参

请求 body 参数:

参数	是否必选	参数类型	示例	描述
evallds	是	[]string	["83"]	评估任务id数组

## 7.13.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	返回的数据对象

data 对象结构:

参数	参数类型	描述
code	string	状态码

参数	参数类型	描述
message	string	状态消息
result	object	任务详细信息
status	int	状态值

### 7.13.7 示例

请求示例:

GET

'https://{endpoint}/api/aidp/alpha/maas\_evaluation\_modelview/api/evaluation/delete?requestInfo=region1'

--header 'Content-Type: application/json'

--data-raw '{"evallds":["83"]}'

正常返回示例:

HTTP/1.1 200 OK

Content-Type:application/json

```
{
  "requestId": "e5b180770f204a89bfede330c75aac61",
  "code": "Success",
  "message": "成功",
  "data": {
    "code": "",
    "message": "success",
    "result": {},
    "status": 0
  }
}
```

## 7.14 大模型评估任务详情接口

### 7.14.1 简要描述

大模型评估任务详情。

### 7.14.2 接口约束

无。



### 7.14.3 URL 格式

https://{endpoint}/api/aidp/alpha/maas\_evaluation\_modelview/api/evaluation/detail?requestInfo={regionId}

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
{regionId}	必须	string	region可用区

### 7.14.4 请求方式

GET。

### 7.14.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type	是	-	application/json	http内容类型, 根据实际情况传参

请求 body 参数:

参数	是否必选	参数类型	示例	描述
id	是	int	68	评估任务id

### 7.14.6 返回参数

参数	参数类型	描述
message	string	信息描述
result	object	任务运行的执行信息
status	int	状态值

result 对象结构:

参数	参数类型	描述
billing_mode	string	计费模式
datasetId	string	数据集ID
datasetName	string	数据集名称
description	string	描述
evalStandardConf	object	评估标准配置

参数	参数类型	描述
evaluationId	integer	评估ID
modelInfoConf	object	模型信息配置
name	string	评估实例名称
projectId	integer	项目ID
resource_id	string	资源ID
serverName	string	服务器名称
service_resource_id	string	服务资源ID
specCPU	string	CPU规格
specCode	string	规格代码
specGpu	string	GPU规格
specMemory	string	内存规格
spec_type	string	规格类型
state	string	状态
storage_type	string	存储类型
task_specCPU	string	任务CPU规格
task_specCode	string	任务规格代码
task_specGpu	string	任务GPU规格
task_specMemory	string	任务内存规格

### 7.14.7 示例

请求示例：

POST

'https://{endpoint}/api/aidp/alpha/maas\_evaluation/api/evaluation/detail?requestInfo=region1'

--header 'Content-Type: application/json'

--data-raw '{"id":"68"}'

正常返回示例：

HTTP/1.1 200 OK

Content-Type:application/json

{

  "message": "success",

  "result": {

    "billing\_mode": "",

    "datasetId": "142",

```
"datasetName": "\u6211\u7684\u6570\u636e\u96c6>hsk-test2",
"description": "",
"evalStandardConf": {
  "evalMode": "manual",
  "evaluationDimension": [
    {
      "description": "",
      "dimension": "\u6ee1\u610f\u5ea6",
      "key": 1
    }
  ],
  "stopWordsPath": ""
},
"evaluationId": 68,
"modelInfoConf": {
  "aihub_id": 41,
  "maas_task_parent_id": "",
  "model_name": "chinese-llama-2-7b",
  "model_path": "",
  "model_select": "\u9884\u7f6e\u6a21\u578b",
  "model_source": "",
  "model_state": "online",
  "model_type": ""
},
"name": "eval-hsk-ch",
"projectId": 48,
"resource_id": "AIDP-36e83f363ff041a2884d5b4175855af8",
"serverName": "",
"service_resource_id": "AIDP-36e83f363ff041a2884d5b4175855af8",
"specCPU": "2",
"specCode": "standard",
"specGpu": "1",
"specMemory": "4G",
"spec_type": "compute",
"state": "Stopped",
"storage_type": "0",
"task_specCPU": "2",
"task_specCode": "standard",
```

```

    "task_specGpu": "1",
    "task_specMemory": "4G"
  },
  "status": 0
}

```

## 7.15 大模型任务创建接口

### 7.15.1 简要描述

大模型任务创建。

### 7.15.2 接口约束

无。

### 7.15.3 URL 格式

大模型任务类型	URL 格式
预训练	https://{endpoint}/api/aidp/alpha/maas_modelview/pt/api
调优SFT	https://{endpoint}/api/aidp/alpha/maas_modelview/sft/api
调优DPO	https://{endpoint}/api/aidp/alpha/maas_modelview/dpo/api
调优RLHF—奖励模型训练	https://{endpoint}/api/aidp/alpha/maas_modelview/rw/api
调优RLHF—强化学习训练	https://{endpoint}/api/aidp/alpha/maas_modelview/ppo/api

URL 参数说明：

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址

### 7.15.4 请求方式

POST。

### 7.15.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type	是	-	application/json	http内容类型，根据实际情况传参

请求 body 参数:

参数	是否必选	参数类型	示例	描述
project	是	int	68	项目id
name	是	string	test-0001	任务名
describe	是	string	测试项目	任务描述

## 7.15.6 返回参数

参数	参数类型	描述
message	string	信息描述
result	object	任务运行的执行信息
status	int	状态值

result 对象结构:

参数	参数类型	描述
describe	string	任务描述
id	int	任务ID
name	string	任务名称
project	string	项目id

## 7.15.7 示例

请求示例: POST [https://cecstackzsarm.user.cecloud.com/api/aidp/alpha/maas\\_modelview/pt/api/](https://cecstackzsarm.user.cecloud.com/api/aidp/alpha/maas_modelview/pt/api/)  
--header 'Content-Type: application/json'  
--data-raw '{"project":49,"name":"test-0001","describe":"test"}

正常返回示例:

HTTP/1.1 200 OK

Content-Type:application/json

```
{  
  "message": "success",  
  "result": {  
    "describe": "test",  
    "id": 74,  
    "name": "test-0001",  
    "project": 49
```

```

    },
    "status": 0
  }

```

## 7.16 大模型任务删除接口

### 7.16.1 简要描述

大模型任务删除。

### 7.16.2 接口约束

无。

### 7.16.3 URL 格式

大模型任务类型	URL 格式
预训练	https://{endpoint}/api/aidp/alpha/maas_modelview/pt/api/{taskId}
调优SFT	https://{endpoint}/api/aidp/alpha/maas_modelview/sft/api/{taskId}
调优DPO	https://{endpoint}/api/aidp/alpha/maas_modelview/dpo/api/{taskId}
调优RLHF—奖励模型训练	https://{endpoint}/api/aidp/alpha/maas_modelview/rw/api/{taskId}
调优RLHF—强化学习训练	https://{endpoint}/api/aidp/alpha/maas_modelview/ppo/api/{taskId}

URL 参数说明：

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
{taskId}	必须	int	任务id

### 7.16.4 请求方式

DELETE。

### 7.16.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type	是	-	application/json	http内容类型，根据实际情况传参

## 7.16.6 返回参数

参数	参数类型	描述
message	string	信息描述
result	object	任务运行的执行信息
status	int	状态值

result 对象结构:

参数	参数类型	描述
alert_status	string	警报状态, 可能的值为 "Pending,Running,Succeeded,Failed,Terminated"
alert_user	string	警报用户, 空字符串表示未设置
changed_by_fk	integer	最后修改用户的外键ID
changed_on	string	最后修改时间, ISO 8601格式
created_by_fk	integer	创建用户的外键ID
created_on	string	创建时间, ISO 8601格式
cron_time	string	定时任务时间, 未设置则为null
cronjob_start_time	string	定时任务开始时间, 格式为"YYYY-MM-DD HH:MM:SS"
dag_json	string	DAG配置的JSON字符串, 空对象表示未设置
depends_on_past	boolean	是否依赖过去任务的完成
describe	string	描述信息, 包含废弃字符
expand	string	扩展信息, 未设置
expired_limit	integer	过期限制, 未设置
global_env	string	全局环境变量, 未设置
id	integer	唯一标识ID
image_pull_policy	string	镜像拉取策略
max_active_runs	integer	最大活跃运行数
name	string	名称
namespace	string	命名空间
node_selector	string	节点选择器
parallelism	integer	并行度
parameter	string	参数, JSON字符串
pipeline_argo_id	int	Argo流水线ID, 未设置则为null
pipeline_file	string	流水线文件路径, 未设置

project_id	integer	项目ID
run_id	int	运行ID, 未设置则为null
schedule_type	string	调度类型, "once"表示一次性
type	string	类型
version_id	int	版本ID, 未设置则为null

## 7.16.7 示例

请求示例:

```
DELETE 'https://{endpoint}/api/aidp/alpha/maas_modelview/pt/api/74?requestInfo=region1'
--header 'Content-Type: application/json'
```

正常返回示例:

HTTP/1.1 200 OK

Content-Type:application/json

```
{
  "message": "success",
  "result": {
    "alert_status": "Pending,Running,Succeeded,Failed,Terminated",
    "alert_user": "",
    "changed_by_fk": 14,
    "changed_on": "2024-07-12T15:39:05",
    "created_by_fk": 14,
    "created_on": "2024-07-12T15:36:45",
    "cron_time": null,
    "cronjob_start_time": "2024-07-12 15:36:45",
    "dag_json": "{}",
    "depends_on_past": false,
    "describe": "test(\u5e9f\u5f03)",
    "expand": "",
    "expired_limit": 0,
    "global_env": "",
    "id": 74,
    "image_pull_policy": "IfNotPresent",
    "max_active_runs": 3,
    "name": "test-0001",
    "namespace": "pipeline",
    "node_selector": "cpu=true,train=true",
```



```
"parallelism": 1,  
"parameter": "{}",  
"pipeline_argo_id": null,  
"pipeline_file": "",  
"project_id": 49,  
"run_id": null,  
"schedule_type": "once",  
"type": "pt",  
"version_id": null  
},  
"status": 0  
}
```

# 8 镜像 API 参考

## 8.1 产品功能简介

智算开发机镜像功能提供了一种快速、一致且可定制的解决方案，用于部署预装有特定软件和工具的虚拟开发环境，从而简化了开发设置，加快了项目启动速度，并确保跨团队成员的环境一致性，同时支持版本控制和安全性更新，以提高开发效率和保障项目安全。

## 8.2 镜像构建接口

### 8.2.1 简要描述

构建镜像。

### 8.2.2 接口约束

无。

### 8.2.3 URL 格式

`https://{endpoint}/api/aidp/v1/docker_modelview/api/`

URL 参数说明：

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为aidp服务
/v1	必须	string	表示接口版本号为：v1
/docker_modelview/api/	必须	string	表示构建镜像

### 8.2.4 请求方式

POST。

### 8.2.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

请求 body 参数:

参数	是否必选	参数类型	描述	示例
project	是	Number	项目ID。	3
image_type	是	String	镜像类型，例如“基础镜像构建”。	
base_image	是	String	基础镜像的地址。	
consecutive_build	是	Boolean	是否连续构建， <b>true</b> 表示是， <b>false</b> 表示否。	true
volume_mount	是	String	挂载点路径。	
describe	是	String	描述信息。	
target_image	是	String	目标镜像的地址。	
resourceId	是	String	资源ID。	
specCode	是	String	规格代码。	
resource_memory	是	String	分配的内存（单位： <b>GB</b> ）。	
resource_cpu	是	String	分配的CPU数量。	
resource_gpu	是	String	分配的GPU数量。	
exclusive	是	String	是否独占资源，若为空表示不独占。	

## 8.2.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	
code	String	返回代码。
message	String	返回消息，表示请求结果。
result	Object	详细结果对象。
├ base_image	String	基础镜像的地址。
├ consecutive_build	Boolean	是否连续构建。
├ describe	String	描述信息。
├ dockerfile	String	Dockerfile内容。
├ expand	String	扩展信息，包含挂载点、资源配置等。

参数	参数类型	描述
└ id	Number	唯一标识ID。
└ image_type	String	镜像类型。
└ project	Number	项目ID。
└ target_image	String	目标镜像的地址。
status	Number	请求状态码，0表示成功。

## 8.2.7 示例

请求示例：

```
curl 'https://cecstackzsarm.user.cecloud.com/api/aidp/v1/docker_modelview/api/' \
-H 'Accept: application/json, text/plain, */*' \
-H 'x-auth-validate: true' \
-H 'x-regionid: region1' \
--data-raw '{"project":36,"image_type":"基础镜像构建","base_image":"image.cestc.cn/iaas_pub/ubuntu-gpu:cuda11.8.0-cudnn8-python3.9","consecutive_build":true,"volume_mount":"/mnt/1841132031866568705","describe":"test","target_image":"cke-public-ccr.registry.region1.cecstackzsarm.cn/library/1841132031866568705:2024.07.12.1","resourceId":"AIDP-c01bb1977eaf4f4dbe0756cfcaa8d7f2","specCode":"standardCpu","resource_memory":"4","resource_cpu":"2","resource_gpu":"0","exclusive":""}' \
--insecure
```

正常返回示例：

HTTP/1.1 200 OK

Content-Type:application/json

```
{
  "requestId": "c6ed66757b3d4da5ade4b5b23e429598",
  "code": "Success",
  "message": "成功",
  "data": {
    "code": "",
    "message": "success",
    "result": {
      "base_image": "image.cestc.cn/iaas_pub/ubuntu-gpu:cuda11.8.0-cudnn8-python3.9",
      "consecutive_build": true,
      "describe": "test",
```

```

        "dockerfile": "",
        "expand": "{\n                \"volume_mount\": \"\/mnt\/1841132031866568705\", \n
\"resource_memory\": \"4G\", \n                \"resource_cpu\": \"2\", \n                \"resource_gpu\": \"0\", \n
\"namespace\": \"jupyter\", \n                \"id\": 3,
        \"image_type\": \"基础镜像构建\",
        \"project\": 36,
        \"target_image\": \"cke-public-ccr.registry.region1.cecstackzsarm.cn\/library\/1841132031866568705:2024.07.12\"
    },
    \"status\": 0
}
}
}

```

## 8.3 镜像构建删除接口

### 8.3.1 简要描述

删除镜像构建。

### 8.3.2 接口约束

无。

### 8.3.3 URL 格式

<https://{endpoint}/api/aidp/v1/deleteInstance>

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/v1	必须	string	表示接口版本号为: v1
/deleteInstance	必须	string	表示删除资源

### 8.3.4 请求方式

DELETE。

## 8.3.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型, 根据实际情况传参

请求 body 参数:

参数	是否必选	参数类型	描述	示例
id	是	int	资源id	3
type	是	int	资源类型, 对于镜像为4	4

## 8.3.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	
code	String	返回代码。
message	String	返回消息, 表示请求结果。
result	Object	详细结果对象。
├ base_image	String	基础镜像的地址。
├ changed_by_fk	Number	最后修改用户的唯一标识。
├ changed_on	String	最后修改时间 (ISO 8601格式)。
├ consecutive_build	Boolean	是否连续构建。
├ created_by_fk	Number	创建用户的唯一标识。
├ created_on	String	创建时间 (ISO 8601格式)。
├ describe	String	描述信息。
├ dockerfile	String	Dockerfile内容。
├ entrypoint	String	入口点命令。
├ expand	String	扩展信息, 包含挂载点、资源配置等。
├ id	Number	唯一标识ID。
├ image_type	String	镜像类型。
├ last_image	String	上次使用的镜像地址。

参数	参数类型	描述
└─ need_gpu	Boolean	是否需要GPU。
└─ project_id	Number	项目ID。
└─ target_image	String	目标镜像的地址。
status	Number	请求状态码，0表示成功。

### 8.3.7 示例

请求示例：

```
curl 'https://cecstackzsarm.user.cecloud.com/api/aidp/v1/deleteInstance?id=3&type=4' \
-X 'DELETE' \
-H 'Accept: application/json, text/plain, */*' \
-H 'x-regionid: region1' \
--insecure
```

正常返回示例：

HTTP/1.1 200 OK

Content-Type:application/json

```
{
  "requestId": "e08cbfb7ba03461dbe307769a9acb967",
  "code": "Success",
  "message": "成功",
  "data": {
    "code": "",
    "message": "success",
    "result": {
      "base_image": "image.cestc.cn/iaas_pub/ubuntu-gpu:cuda11.8.0-cudnn8-python3.9",
      "changed_by_fk": 12,
      "changed_on": "2024-07-12T11:20:54",
      "consecutive_build": true,
      "created_by_fk": 12,
      "created_on": "2024-07-12T11:20:54",
      "describe": "test",
      "dockerfile": "",
      "entrypoint": "",
      "expand": "{\n          \"volume_mount\": \"\n          \"/mnt/1841132031866568705\", \n          \n          \"resource_memory\": \"4G\", \n          \"resource_cpu\": \"2\", \n          \"resource_gpu\": \"0\", \n          \"namespace\": \"jupyter\" \n        }",
```

```

    "id": 3,
    "image_type": "基础镜像构建",
    "last_image": "",
    "need_gpu": false,
    "project_id": 36,
    "target_image": "cke-public-
ccr.registry.region1.cecstackzsarm.cn/library/1841132031866568705:2024.07.12.1"
  },
  "status": 0
}
}

```

## 8.4 镜像列表查询接口

### 8.4.1 简要描述

查询镜像列表。

### 8.4.2 接口约束

无。

### 8.4.3 URL 格式

[https://{endpoint}/api/aidp/alpha/images\\_modelview/api/](https://{endpoint}/api/aidp/alpha/images_modelview/api/)

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/alpha	必须	string	表示接口版本号为: alpha
/images_modelview/api/	必须	string	表示镜像列表查询接口

### 8.4.4 请求方式

GET。

### 8.4.5 请求参数

请求 header 参数:

无



请求 body 参数:

参数	是否必选	参数类型	描述	示例
str_related	是	String	相关字符串。	"1"
page	是	Number	页码, 默认为0。	0
page_size	是	Number	每页条目数。	10
filters	是	Array	过滤条件数组。	
├ col	是	String	过滤列名。	"isPreset"
├ opr	是	String	操作符, 例如"eq"表示等于。	"eq"
├ value	是	Any	过滤值。	true

## 8.4.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	
creator	String	创建者用户名。
gpu_type	String	GPU类型。
gpu_type_ch	String	GPU类型中文描述
id	Number	唯一标识ID。
image_type	String	镜像类型。
images_url	String	镜像URL地址
isPreset	Boolean	是否预设
key	String	镜像键值。
modified	String	最后修改时间。
project	String	项目名称。
title	String	标题。

## 8.4.7 示例

请求示例:

curl

'https://cecstackzsarm.user.cecloud.com/api/aidp/alpha/images\_modelview/api/?form\_data=%7B

```
%22str_related%22:%221%22,%22page%22:0,%22page_size%22:10,%22filters%22:\[%7B%22col%22:%22isPreset%22,%22opr%22:%22eq%22,%22value%22:true%7D\]%7D' \
```

```
-H 'Accept: application/json, text/plain, */*' \
```

```
-H 'Accept-Language: en-US,en;q=0.9,zh;q=0.8' \
```

```
-H 'Cache-Control: no-cache' \
```

```
-H 'Connection: keep-alive' \
```

```
-H 'Cookie: cid_raoyunkun_cestc.cn=TUhJeVIUQnZPWGsyZFRadU9HczVkVEp1;_ga=GA1.2.642436277.1719401511;
```

```
_ga_R1FN4KJKJH=GS1.1.1719401511.1.1.1719401912.0.0.0;
```

```
ea6ee7ef65afa3fa0312817d0b12190e=c81242f20bef883c2491198f5f4cc241;
```

```
myapp_username=1841132031866568705;
```

```
CECLOUD_OPS_CID=TkdReFkySmtaR1F0TkdZeVITMDBNbU5rTFRoa016RZRZamN5TIRJek9UTXIOV1k1fFkyVmpjM1JoWTJ0NmMyRnliUzUxYzJWeUxtTmxZMnh2ZFdRdVkyOXQ=;
```

```
CECLOUD_OPS_SID=TkRkbU1tVXpZekl0WkRjd05pMDBNelZoTFdKaFlqQXRPREI6TTJJelpURmpaV05ofFkyVmpjM1JoWTJ0NmMyRnliUzUxYzJWeUxtTmxZMnh2ZFdRdVkyOXQ=;
```

```
session=eyJfZmxhc2hlcyI6W10sIl9mcmVzaCI6ZmFsc2V9.ZpC2jw.4e8-
```

```
_LppWkSNveuTqL_rBqvWrh4' \
```

```
-H 'Pragma: no-cache' \
```

```
-H
```

```
'Referer:
```

```
https://cecstackzsarm.user.cecloud.com/region/region1/console/tenant/compute/aidp/image-manage/mirror-list' \
```

```
-H 'x-regionid: region1' \
```

```
--insecure
```

正常返回示例:

```
HTTP/1.1 200 OK
```

```
Content-Type:application/json
```

```
{
```

```
  "message": "success",
```

```
  "result": {
```

```
    "count": 370,
```

```
    "data": [
```

```
      {
```

```
        "creator": "admin",
```

```
        "gpu_type": "ascend",
```

```
        "gpu_type_ch": "\u534e\u4e3a\u6607\u817e",
```

```
        "id": 373,
```

```
        "image_type": "",
```

```
        "images_url": "image.cestc.cn/iaas_pub/ascend-llm-trainer:v0.1-20240529",
```

```
        "isPreset": true,
```

```

    "key": "ascend-llm-trainer:v0.1-20240529",
    "modified": "15 days ago",
    "project": "\u5927\u6a21\u578b",
    "title":
"\u5927\u6a21\u578b\u670d\u52a1\u6607\u817eGPU\u7248\u8bad\u7ec3\u529fu80
fd\u09"
  },
  {
    "creator": "admin",
    "gpu_type": "ascend",
    "gpu_type_ch": "\u534e\u4e3a\u6607\u817e",
    "id": 372,
    "image_type": "",
    "images_url": "image.cestc.cn/iaas_pub/ascend-llm-server:v0.1-20240529",
    "isPreset": true,
    "key": "ascend-llm-server:v0.1-20240529",
    "modified": "15 days ago",
    "project": "\u5927\u6a21\u578b",
    "title":
"\u5927\u6a21\u578b\u670d\u52a1\u6607\u817eGPU\u7248\u63a8\u7406\u529fu80
fd\u09"
  },
  {
    "creator": "admin",
    "gpu_type": "iluvatar_infer",
    "gpu_type_ch": "\u5929\u6570\u667a\u82af\u63a8\u7406\u09",
    "id": 371,
    "image_type": "",
    "images_url": "image.cestc.cn/iaas_pub/iluvatar-llm-server:v0.1-20240624",
    "isPreset": true,
    "key": "iluvatar-llm-server:v0.1-20240624",
    "modified": "15 days ago",
    "project": "\u5927\u6a21\u578b",
    "title":
"\u5927\u6a21\u578b\u670d\u52a1\u6570GPU\u7248\u63a8\u7406\u529fu80
fd\u09"
  },
  {
    "creator": "admin",

```

```

    "gpu_type": "nvidia",
    "gpu_type_ch": "\u82f1\u4f1f\u8fbe",
    "id": 370,
    "image_type": "",
    "images_url": "image.cestc.cn/iaas_pub/opencompass:v2.2.2",
    "isPreset": true,
    "key": "opencompass:v2.2.2",
    "modified": "16 days ago",
    "project": "\u5927\u6a21\u578b",
    "title": "\u5927\u6a21\u578b\u57fa\u7840\u8bc4\u4f30"
  },
  {
    "creator": "admin",
    "gpu_type": "nvidia",
    "gpu_type_ch": "\u82f1\u4f1f\u8fbe",
    "id": 369,
    "image_type": "",
    "images_url": "image.cestc.cn/iaas_pub/model-evaluation:v1.0-20240621-48375",
    "isPreset": true,
    "key": "model-evaluation:v1.0-20240621-48375",
    "modified": "17 days ago",
    "project": "\u5927\u6a21\u578b",
    "title": "\u5927\u6a21\u578b\u8bc4\u4f30"
  },
  {
    "creator": "admin",
    "gpu_type": "nvidia",
    "gpu_type_ch": "\u82f1\u4f1f\u8fbe",
    "id": 368,
    "image_type": "",
    "images_url": "image.cestc.cn/ccos-aidp/preset-models:SUS-Chat-34B",
    "isPreset": true,
    "key": "preset-models:SUS-Chat-34B",
    "modified": "17 days ago",
    "project": "\u9884\u5236\u6a21\u578b",
    "title": "\u9884\u5236\u5927\u6a21\u578bSUS-Chat-34B"
  },
  {

```

```

    "creator": "admin",
    "gpu_type": "nvidia",
    "gpu_type_ch": "\u82f1\u4f1f\u8fbc",
    "id": 367,
    "image_type": "",
    "images_url": "image.cestc.cn/ccos-aidp/preset-models:zephyr-7b-beta",
    "isPreset": true,
    "key": "preset-models:zephyr-7b-beta",
    "modified": "17 days ago",
    "project": "\u9884\u5236\u6a21\u578b",
    "title": "\u9884\u5236\u5927\u6a21\u578bzephyr-7b-beta"
  },
  {
    "creator": "admin",
    "gpu_type": "nvidia",
    "gpu_type_ch": "\u82f1\u4f1f\u8fbc",
    "id": 366,
    "image_type": "",
    "images_url": "image.cestc.cn/ccos-aidp/preset-models:Yi-34B-Chat",
    "isPreset": true,
    "key": "preset-models:Yi-34B-Chat",
    "modified": "17 days ago",
    "project": "\u9884\u5236\u6a21\u578b",
    "title": "\u9884\u5236\u5927\u6a21\u578bYi-34B-Chat"
  },
  {
    "creator": "admin",
    "gpu_type": "nvidia",
    "gpu_type_ch": "\u82f1\u4f1f\u8fbc",
    "id": 365,
    "image_type": "",
    "images_url": "image.cestc.cn/ccos-aidp/preset-models:Ziya2-13B-Chat",
    "isPreset": true,
    "key": "preset-models:Ziya2-13B-Chat",
    "modified": "17 days ago",
    "project": "\u9884\u5236\u6a21\u578b",
    "title": "\u9884\u5236\u5927\u6a21\u578bZiya2-13B-Chat"
  },
},

```

```

    {
      "creator": "admin",
      "gpu_type": "nvidia",
      "gpu_type_ch": "\u82f1\u4f1f\u8fbe",
      "id": 364,
      "image_type": "",
      "images_url": "image.cestc.cn/iaas_pub/llm-server:1.0.0-5afef2c-20240607-1524-gpu",
      "isPreset": true,
      "key": "llm-server:1.0.0-5afef2c-20240607-1524-gpu",
      "modified": "17 days ago",
      "project": "\u5927\u6a21\u578b",
      "title": "\u5927\u6a21\u578b\u670d\u52a1"
    }
  ],
  "description_columns": {
    "entrypoint":
"\u955c\u50cf\u7684\u5165\u53e3\u547d\u4ee4\u76f4\u63a5\u5199\u6210\u5355\u884c\u5b57\u7b26\u4e32\u4f8b\u5982python xx.py\u65e0\u9700\u6dfb\u52a0[]",
    "gpu_type": "GPU\u7c7b\u578b",
    "name":
"\u955c\u50cf\u540d\u79f0\u5168\u79f0\u4f8b\u5982image.cestc.cn/iaas_pub/ubuntu:20.04"
  },
  "label_columns": {
    "changed_by": "\u4fee\u6539\u8005",
    "changed_on": "\u4fee\u6539\u65f6\u95f4",
    "created_by": "\u521b\u5efa\u8005",
    "created_on": "\u521b\u5efa\u65f6\u95f4",
    "creator": "\u521b\u5efa\u8005",
    "describe": "\u63cf\u8ff0",
    "dockerfile": "Dockerfile",
    "entrypoint": "\u542f\u52a8\u547d\u4ee4",
    "gitpath": "git\u5730\u5740",
    "gpu_type": "GPU\u7c7b\u578b",
    "gpu_type_ch": "GPU\u7c7b\u578b",
    "image_type": "\u955c\u50cf\u7c7b\u578b",
    "images_url": "\u955c\u50cf",
    "isPreset": "\u662f\u5426\u9884\u5236",
  }

```

```

    "key": "\u955c\u50cf\u6807\u8bc6",
    "modified": "\u4fee\u6539\u65f6\u95f4",
    "name": "\u540d\u79f0",
    "project": "\u529f\u80fd\u5206\u7c7b",
    "repository": "\u4ed3\u5e93",
    "title": "\u955c\u50cf\u6807\u9898",
    "usage_type": "\u4f7f\u7528\u573a\u666f"
  },
  "list_columns": [
    "project",
    "title",
    "images_url",
    "key",
    "creator",
    "modified",
    "image_type",
    "gpu_type",
    "isPreset",
    "gpu_type_ch"
  ],
  "list_title": "\u955c\u50cf \u5217\u8868",
  "order_columns": [
    "id"
  ]
},
"status": 0
}

```

## 8.5 自定义镜像上传接口

### 8.5.1 简要描述

自定义镜像上传分为管理面和数据面 2 部分，用户在前端页面点击上传之后，会同时触发数据面的镜像包上传，相关接口（</api/ccr/v1.0/files/uploads/>）请参考 CCR 接口文档。在数据面完成镜像上传之后，会调用管理面接口完成镜像信息的注册。下面介绍智算平台管理面部分镜像上传接口。

### 8.5.2 接口约束

无。

### 8.5.3 URL 格式

https://{endpoint}/api/aidp/v1/images/preUpload

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/v1	必须	string	表示接口版本号为: v1
/images/preUpload	必须	string	参数校验

### 8.5.4 请求方式

POST。

### 8.5.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型, 根据实际情况传参

请求 body 参数:

参数	是否必选	参数类型	描述	示例
projectName	是	string	项目名称	
imageName	是	string	镜像名称	
abortController	是	object	终止控制器	传 {} 即可

### 8.5.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	
instanceId	string	实例id
projectName	string	仓库名称



## 8.5.7 示例

请求示例:

```
POST 'https://cecstackzsarm.user.cecloud.com/api/aidp/v1/images/preUpload' \  
-H 'Accept: application/json, text/plain, */*' \  
-H 'x-regionid: region1' \  
--data-raw '{"projectName":"library","imageName":"test:v1","abortController":{}}' \  
--insecure
```

正常返回示例:

```
HTTP/1.1 200 OK  
Content-Type:application/json  
{  
  "requestId": "f14c4142fbc54af5988b80a201292f7c",  
  "code": "Success",  
  "message": "成功",  
  "data": {  
    "instanceId": "ccr-unws3g1ctn",  
    "projectName": "library"  
  }  
}
```

## 8.6 镜像推送接口

### 8.6.1 简要描述

镜像推送。

### 8.6.2 接口约束

无。

### 8.6.3 URL 格式

https://{endpoint}/api/aidp/v1/instances/{ccr\_instance\_id}/projects/{project\_name}/library/push

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式

参数	是否必选	参数类型	描述
/aidp	必须	string	表示为aidp服务
/v1	必须	string	表示接口版本号为： v1
/instances/{ccr_instance_id}/projects/{project_name}/push	必须	string	推送镜像

## 8.6.4 请求方式

POST。

## 8.6.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

请求 body 参数：

参数	是否必选	参数类型	描述	示例
hash_md5	是	string	镜像的MD5哈希值	
image_name	是	string	镜像名称	"test:v1"
upload_id	是	string	上传id	"77e9f31cab29348f4ed0482be7aa4f76"

## 8.6.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	
status	string	推送状态

## 8.6.7 示例

请求示例：

POST `'https://cecstackzsarm.user.cecloud.com/api/aidp/v1/instances/ccr-unws3g1ctn/projects/library/push'` \

```
-H 'Accept: application/json, text/plain, */*' \  
-H 'x-regionid: region1' \  
--data-raw  
'{"hash_md5":"","image_name":"test:v1","upload_id":"77e9f31cab29348f4ed0482be7aa4f76"}' \  
--insecure
```

正常返回示例:

HTTP/1.1 200 OK

Content-Type:application/json

```
{  
  "requestId": "5c469d0f1ad144839b7129212ee75fde",  
  "code": "Success",  
  "message": "成功",  
  "data": {  
    "status": "Pushing"  
  }  
}
```

# 9 项目空间参考

## 9.1 项目空间创建接口

### 9.1.1 简要描述

创建项目空间。

### 9.1.2 接口约束

无。

### 9.1.3 URL 格式

URL 参数说明：

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址，此处为本地地址 127.0.0.1
/api	必须	string	统一格式
/aidp	必须	string	表示adip服务
/v1	必须	string	接口版本号为：v1
/project/insert	必须	string	表示项目创建操作
requestInfo	必须	string	请求的区域标识，此处为zsregion0

### 9.1.4 请求方式

POST。

### 9.1.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

请求 body 参数：

参数	是否必选	参数类型	示例	描述
name	是	string	test_namespace	项目名称
Describe	否	string	测试	项目描述

## 9.1.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	boolean	操作结果，true表示成功

## 9.1.7 示例

请求示例：

```
POST 'https://127.0.0.1/api/aidp/v1/project/insert?requestInfo=zsregion0' --header 'Content-Type: application/json' --data-raw '{"name":"test_namespace","Describe":"测试"}
```

正常返回示例：

```
HTTP/1.1 200 OK Content-Type: application/json { "requestId": "bc448c2fa80c40f0a5c138de2f02e6aa", "code": "Success", "message": "成功", "data": true }
```

## 9.2 关联资源接口

### 9.2.1 简要描述

项目空间关联资源。

### 9.2.2 接口约束

无。

### 9.2.3 URL 格式

URL 参数说明：

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址，此处为本地地址 127.0.0.1
/api	必须	string	统一格式

参数	是否必选	参数类型	描述
/aidp	必须	string	表示adip服务
/v1	必须	string	接口版本号为: v1
/resource/project/related	必须	string	表示项目资源关联查询操作
requestInfo	必须	string	请求的区域标识, 此处为zsregion0

## 9.2.4 请求方式

POST。

## 9.2.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型, 根据实际情况传参

请求 body 参数:

参数	是否必选	参数类型	示例	描述
projectId	是	string	"41"	项目ID
resourceIds	是	array	["AIDP-dc81a69afda447f187cb98d4cf486967"]	资源ID列表

## 9.2.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	number	查询到的关联资源数量

## 9.2.7 示例

请求示例:

```
POST 'https://127.0.0.1/api/aidp/v1/resource/project/related?requestInfo=zsregion0' --header
'Content-Type: application/json' --data-raw '{"projectId": "41", "resourceIds": ["AIDP-
dc81a69afda447f187cb98d4cf486967"]}'
```

正常返回示例:

```
HTTP/1.1 200 OK Content-Type: application/json { "requestId":  
"a15d36e536a947caa002ba8810f807e5", "code": "Success", "message": "成功", "data": 1 }
```

## 9.3 删除项目接口

### 9.3.1 简要描述

删除项目。

### 9.3.2 接口约束

无。

### 9.3.3 URL 格式

<https://zstest.user.cecloud.com/api/aidp/v1/project/delete>

### 9.3.4 请求方式

DELETE。

### 9.3.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型, 根据实际情况传参

请求 body 参数:

参数	是否必选	参数类型	示例	描述
requestInfo	必须	string	请求的区域标识, 此处为zsregion0	requestInfo
id	必须	number	要删除的项目ID	id

### 9.3.6 返回参数

参数	参数类型	描述
requestId	string	请求ID

参数	参数类型	描述
code	string	返回码
message	string	请求结果描述
data	object	返回的数据对象，成功时可能包含更多信息，失败时为null

### 9.3.7 示例

请求示例：

DELETE

'https://zstest.user.cecloud.com/api/aidp/v1/project/delete?requestInfo=zsregion0&id=41'

正常返回示例：

HTTP/1.1 500 Internal Server Error

Content-Type: application/json

```
{
  "requestId": "e1d9c2c7719943fbb1d2b0c00a3a9bbc",
  "code": "Success",
  "message": "成功",
  "data": true
}
```

## 9.4 移除资源

### 9.4.1 简要描述

移除项目空间资源绑定。

### 9.4.2 接口约束

无。

### 9.4.3 URL 格式

https://zstest.user.cecloud.com/api/aidp/v1/resource/project/unRelated?requestInfo=zsregion0

URL 参数说明：



参数	是否必选	参数类型	描述
requestInfo	必须	string	请求的区域标识，此处为zsregion0

## 9.4.4 请求方式

POST。

## 9.4.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

请求 body 参数：

参数	是否必选	参数类型	示例	描述
projectId	是	string	"41"	项目ID
resourceIds	是	array	["AIDP-dc81a69afda447f187cb98d4cf486967"]	要解绑的资源ID列表

## 9.4.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	number	解绑的资源数量

## 9.4.7 示例

请求示例：

POST

```
'https://zstest.user.cecloud.com/api/aidp/v1/resource/project/unRelated?requestInfo=zsregion0'
```

```
--header 'Content-Type: application/json'
```

```
--data-raw '{"projectId":"41","resourceIds":["AIDP-dc81a69afda447f187cb98d4cf486967"]}'正常返回示例：
```

HTTP/1.1 200 OK

Content-Type: application/json

```
{
  "requestId": "6d132ff40ea2486fb14c4594c7be5b04",
  "code": "Success",
  "message": "成功",
  "data": 1
}
```

## 9.5 查询项目

### 9.5.1 简要描述

查询项目空间。

### 9.5.2 接口约束

无。

### 9.5.3 URL 格式

<https://zstest.user.cecloud.com/api/aidp/v1/project/page>

### 9.5.4 请求方式

GET。

### 9.5.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

请求 body 参数:

参数	是否必选	参数类型	示例	描述
form_data	是	String	{"filters":[],"str_relate":1,"page":0,"page_size":10}	JSON格式的查询参数

## 9.5.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	分页数据对象
- count	number	总条目数量
- data	array	项目列表

项目列表对象属性:

属性名	参数类型	描述
id	number	项目ID
cluster	string	集群
name	string	项目名称
projectUser	string	项目用户
createTime	string	创建时间
status	string	项目状态
completeResource	string	完整资源名称
storageResource	string	存储资源名称
exclusiveResource	string	独占资源名称

## 9.5.7 示例

请求示例:

```
GET 'https://zstest.user.cecloud.com/api/aidp/v1/project/page'  
--header 'Content-Type: application/json'  
--data-raw '{"filters":[],"str_relate":1,"page":0,"page_size":10}'
```

正常返回示例:

```
HTTP/1.1 200 OK  
Content-Type: application/json  
{  
  "requestId": "20e965fe8eed4a6caa226ec413c2ab54",
```

```
"code": "Success",
"message": "成功",
"data": {
  "count": 4,
  "data": [
    {
      "id": 41,
      "name": "test_namespace",
      "projectUser": "user@admin",
      "createTime": "2024-07-12 11:58:40",
      "status": "使用中",
      "completeResource": "useradmin-2705",
      "storageResource": "",
      "exclusiveResource": ""
    },
    // ... 其他项目数据
  ]
}
}
```

# 10 数据集管理

## 10.1 数据集创建

数据集创建。

### 10.1.1 接口约束

无。

### 10.1.2 URL 格式

<https://cecstackzsarm.user.cecloud.com/api/aidp/v1/dataset/insert>

### 10.1.3 请求方式

POST。

### 10.1.4 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

请求 body 参数:

参数	是否必选	参数类型	描述
name	是	string	aa11
describe	是	string	dddd
label	否	string	
sourceType	是	string	text
source	否	string	
status	否	string	
downloadUrl	是	string	/mnt/59/aa11/v2
markStatus	是	string	2
clearStatus	否	string	
projectId	是	number	59

参数	是否必选	参数类型	描述
version	是	string	v2
storageType	是	number	0

### 10.1.5 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	boolean	操作结果，true表示成功

### 10.1.6 示例

请求示例：

```
POST 'https://cecstackzsarm.user.cecloud.com/api/aidp/v1/dataset/insert'
--header 'Content-Type: application/json'
--data-raw
'{"name":"aa11","describe":"dddd","label":"","sourceType":"text","source":"","status":"","download
Url":"/mnt/59/aa11/v2","markStatus":"2","clearStatus":"","projectId":59,"version":"v2","storageType
":0}'
```

正常返回示例：

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

```
{
  "requestId": "fb122a9315714ee28d1ac29a53210f1e",
  "code": "Success",
  "message": "成功",
  "data": true
}
```

## 10.2 上传文件

### 10.2.1 简要描述

上传数据集文件。

### 10.2.2 接口约束

无。

### 10.2.3 URL 格式

https://zstest.user.cecloud.com/api/aidp/v1/dataset/files

### 10.2.4 请求方式

POST。

### 10.2.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

请求 body 参数：

参数	是否必选	参数类型	示例	描述
namespace	是	string	ccos-aidp	命名空间
dstpath	是	string	/mnt/59/aa11/v2/oa3	文件目标路径
uploadId	是	string	a2ff35d0c3d4b4852c7dea9dda166176	文件上传ID
filename	是	string	oaast_sft_zh.json	文件名称

### 10.2.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	boolean	操作结果，true表示成功

## 10.2.7 示例

请求示例：

```
POST 'https://zstest.user.cecloud.com/api/aidp/v1/dataset/files'  
--header 'Content-Type: application/json'  
--data-raw '{"namespace":"ccos-aidp","dstpath":"/mnt/59/aa11/v2/oa3","uploadId":"a2ff35d0c3d4b4852c7dea9dda166176","filename":"oaast_sft_zh.json"}'
```

正常返回示例：

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

```
{  
  "requestId": "7b206625cb4b4117925532e1595119a9",  
  "code": "Success",  
  "message": "成功",  
  "data": true  
}
```

## 10.3 查询数据集

### 10.3.1 简要描述

数据集查看。

### 10.3.2 接口约束

无。

### 10.3.3 URL 格式

<https://cecstackzsarm.user.cecloud.com/api/aidp/v1/dataset/page>

### 10.3.4 请求方式

GET。



## 10.3.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型, 根据实际情况传参

请求 body 参数:

参数	是否必选	参数类型	示例	描述
filters	否	array	[]	过滤条件数组
str_relate	是	number	1	关联资源类型
page	是	number	0	当前页码
page_size	是	number	10	每页条目数量

## 10.3.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	分页数据对象
- count	number	总条目数量
- data	array	数据集列表

属性名	参数类型	描述
id	number	数据集ID
name	string	数据集名称
describe	string	数据集描述
version	string	版本号
storageSize	string	存储大小
dataSize	string	数据条目数量
sourceType	string	数据源类型
dealStatus	string	数据处理状态
importStatus	number	数据导入状态
markTemplate	string	数据标注模板

markStatus	string	数据标注状态
createTime	string	创建时间
projectID	number	项目ID
projectName	string	项目名称
storageType	number	存储类型
storageName	string	存储名称
isPublic	number	是否公共数据集
createdBy	string	创建者
changedBy	string	最后修改者

### 10.3.7 示例

请求示例:

```
GET 'https://cecstackzsarm.user.cecloud.com/api/aidp/v1/dataset'
--header 'Content-Type: application/json'
--data-raw '{"filters": [], "str_relate": 1, "page": 0, "page_size": 10}'
```

正常返回示例:

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

```
{
  "requestId": "1b5078d79fdf46d8bd53f2003f6407d7",
  "code": "Success",
  "message": "成功",
  "data": {
    "count": 2,
    "data": [
      // 数据集对象列表
    ]
  }
}
```

## 10.4 删除数据集

### 10.4.1 简要描述

删除数据集。

### 10.4.2 接口约束

无。

### 10.4.3 URL 格式

<https://cecstackzsarm.user.cecloud.com/api/aidp/v1/dataset/delete>

URL 参数说明：无

### 10.4.4 请求方式

DELETE。

### 10.4.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

请求 body 参数：

参数	是否必选	参数类型	示例	描述
id	是	number	150	要删除的数据集ID

### 10.4.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	boolean	操作结果，true表示成功

## 10.4.7 示例

请求示例:

```
DELETE 'https://cecstackzsarm.user.cecloud.com/api/aidp/v1/dataset/delete'  
--header 'Content-Type: application/json'  
--data-raw '{"id": 150}'
```

正常返回示例:

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

```
{  
  "requestId": "d7e66265ec6840f6911073fd3eb8d104",  
  "code": "Success",  
  "message": "成功",  
  "data": true  
}
```

# 11 训练作业 API 参考

## 11.1 产品功能简介

训练作业是 AI 开发平台为机器学习开发者提供的运行代码和模型开发的模块。训练作业支持调试和运行用户代码用于模型训练和调试，也提供了持久化的对象存储服务和共享文件系统用于存储开发中的数据。通过训练作业可以完成用户的代码一次执行并保存执行结果和日志，执行完成即释放用户分配的运行资源，提高资源分配和利用效率。

## 11.2 训练作业创建接口

### 11.2.1 简要描述

创建训练作业任务。

### 11.2.2 接口约束

无。

### 11.2.3 URL 格式

`https://{endpoint}/api/aidp/alpha/maas_task_modelview/slm/api/`

URL 参数说明：

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/alpha	必须	string	表示接口版本号为：alpha
/slm/api/	必须	string	表示为训练作业任务

### 11.2.4 请求方式

POST。

### 11.2.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

请求 body 参数:

参数	参数类型	是否必选	描述	示例
projectId	int	是	项目ID	58
project_id	int	是	项目ID	58
name	string	是	训练作业任务名称	test-train
maas_task_type	string	是	任务类型	slm
billing_mode	string	是	计费模式	resourcePool
label	string	否	训练作业任务标签	test
images	string	是	镜像地址	image.ccos.io/iaas_pub/notebook:vscode-gpu
command	string	是	启动脚本	sleep 30
spec_type	string	是	规格类型	compute
resourceId	string	是	资源ID	AIDP-ffa8f37b95a24b439de3966499a1b3e6
node_selector	string	是	节点选择	train=true
expand	string	是	JSON格式的扩展参数：镜像类型和环境变量	{"trainWorkArgs":{"imageType":"preset"},"envs":[{"key":"KEY","value":"Value"}]}

## 11.2.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	
- code	string	用户id
- message	string	用户名
- result	object	训练作业相关信息
- status	string	状态码

## 11.2.7 示例

请求示例:

```
POST 'https://{endpoint}/api/aidp/alpha/maas_task_modelview/slm/api/'
--header 'Content-Type: application/json'
--data-raw
{"maas_task_type":"slm","billing_mode":"resourcePool","name":"test-
train","label":"label","images":"image.ccos.io/iaas_pub/notebook:jupyter-cpu","command":"sleep
60","spec_type":"compute","resource_id":"AIDP-
dabd0dc99a9447d3980b62c73b0078dd","node_selector":"train=true","projectId":58,"project_id":5
8,"expand":{"trainWorkArgs":{"imageType":"preset"},"envs":{"key":"KEY","value":"Value
\}}}}
```

正常返回示例:

```
{
  "requestId": "42d68d1f2f624dcb867d5e0eab25dd5c",
  "code": "Success",
  "message": "成功",
  "data": {
    "code": "",
    "message": "success",
    "result": {
      "advanced_mode": "close",
      "args": "{}",
      "batch_size": 4,
      "billing_mode": "resourcePool",
      "command": "sleep 60",
      "dataset_id": null,
      "dataset_ids": null,
      "default_system": null,
      "epoch": 3,
      "exclusive": "",
      "expand":
      "{\n\"trainWorkArgs\":{\n\"envs\":{\n\"key\":\n\"KEY\", \"value\":\n\"Value\"}},\n\"imageType\":\n\"preset\"}}",
      "framework": "deepspeed",
      "framework_config": null,
      "framework_template": "",
      "ft_type": "",

```

```

    "id": 5,
    "images": "image.ccos.io/iaas_pub/notebook:jupyter-cpu",
    "label": "label",
    "level": "L0",
    "maas_task_type": "slm",
    "model_path": "",
    "model_type": "",
    "name": "test-train",
    "node_selector": "train=true",
    "output_dir": "",
    "project_id": 58,
    "resource_cpu": "2",
    "resource_gpu": "0",
    "resource_id": "AIDP-dabd0dc99a9447d3980b62c73b0078dd",
    "resource_memory": "4G",
    "retry": 0,
    "spec_type": "compute",
    "storage_type": null,
    "volume_mount": "kubeflow-user-workspace(pvc):/mnt",
    "workers": 1
  },
  "status": 0
}
}

```

## 11.3 训练作业编辑接口

### 11.3.1 简要描述

编辑训练作业任务。

### 11.3.2 接口约束

无。

### 11.3.3 URL 格式

[https://{endpoint}/api/aidp/alpha/maas\\_task\\_modelview/slm/api/{train\\_work\\_id}](https://{endpoint}/api/aidp/alpha/maas_task_modelview/slm/api/{train_work_id})

URL 参数说明:



参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/alpha	必须	string	表示接口版本号为: alpha
/slm/api/	必须	string	表示为训练作业任务
{train_work_id}	必须	Int	表示训练作业ID

### 11.3.4 请求方式

PUT。

### 11.3.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型, 根据实际情况传参

请求 body 参数:

参数	参数类型	是否必选	描述	示例
projectId	int	是	项目ID	58
project_id	int	是	项目ID	58
name	string	是	训练作业任务名称	test-train
maas_task_type	string	是	任务类型	slm
billing_mode	string	是	计费模式	resourcePool
label	string	否	训练作业任务标签	test
images	string	是	镜像地址	image.ccos.io/iaas_pub/notebook:vscode-gpu
command	string	是	启动脚本	sleep 300
spec_type	string	是	规格类型	compute
resourceId	string	是	资源ID	AIDP-ffa8f37b95a24b439de3966499a1b3e6
node_selector	string	是	节点选择	train=true

参数	参数类型	是否必选	描述	示例
expand	string	是	JSON格式的扩展参数：镜像类型和环境变量	{\trainWorkArgs\:{\imageType\:\preset\,\envs\:{\key\:\KEY\,\value\:\Value1\}}}

### 11.3.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	
- code	string	用户id
- message	string	用户名
- result	object	训练作业相关信息
- status	string	状态码

### 11.3.7 示例

请求示例：

```
PUT 'https://{endpoint}/api/aidp/alpha/maas_task_modelview/slm/api/5'
--header 'Content-Type: application/json'
--data-raw
{"maas_task_type":"slm","billing_mode":"resourcePool","name":"test-
train","label":"label","images":"image.ccos.io/iaas_pub/notebook:jupyter-cpu","command":"sleep
600","spec_type":"compute","resource_id":"AIDP-
dabd0dc99a9447d3980b62c73b0078dd","node_selector":"train=true","projectId":58,"project_id":5
8,"expand":{\trainWorkArgs\:{\imageType\:\preset\,\envs\:{\key\:\KEY\,\value\:\Value
\}}}}
```

正常返回示例：

```
{
  "requestId": "ae65762330524ea6b8168039c3612830",
  "code": "Success",
  "message": "成功",
  "data": {
    "code": "",
    "message": "success",
```

```

"result": {
  "advanced_mode": "close",
  "args": "{}",
  "batch_size": 4,
  "billing_mode": "resourcePool",
  "changed_by": 4,
  "changed_on": "2025-01-20T17:21:45",
  "command": "sleep 600",
  "created_by": 4,
  "created_on": "2025-01-20T16:57:38",
  "creator_first_name": "guodx1",
  "dataset": "",
  "dataset_id": null,
  "dataset_ids": null,
  "default_system": null,
  "elapsed_time": "未知",
  "epoch": 3,
  "exclusive": "",
  "expand": "{\n  \"trainWorkArgs\": {\n    \"envs\": [\n      {\n        \"key\": \"KEY\",
\n\"value\": \"Value\"\n      }\n    ],\n    \"imageType\": \"preset\"\n  }\n}",
  "final_status": "\u003csvg t=\"1669360410529\" class=\"icon\" viewBox=\"0 0 1024 1024\"
version=\"1.1\" xmlns=\"http://www.w3.org/2000/svg\" p-id=\"6711\" width=\"20\"
height=\"20\" \u003e\u003cpath d=\"M937.984 741.376c-6.144 10.24-18.432 14.336-28.672
8.192-10.24-6.144-14.336-18.432-8.192-28.672 118.784-212.992 40.96-481.28-172.032-
598.016-212.992-118.784-481.28-40.96-598.016 172.032s-40.96 481.28 172.032 598.016c153.6
86.016 339.968 69.632 479.232-32.768 8.192-6.144 22.528-4.096 28.672 4.096 6.144 8.192
4.096 22.528-4.096 28.672-151.552 112.64-356.352 129.024-522.24 36.864-233.472-129.024-
317.44-421.888-188.416-653.312 129.024-233.472 421.888-317.44 653.312-188.416 233.472
126.976 317.44 419.84 188.416 653.312z m-647.168-243.712i190.464 169.984 282.624-
303.104c8.192-8.192 20.48-8.192 28.672 0 8.192 8.192 8.192 20.48 0 28.672i-311.296 331.776-
219.136-198.656c-8.192-8.192-8.192-20.48-2.048-28.672 8.192-8.192 20.48-8.192 30.72 0z\" p-
id=\"6712\" fill=\"#dbdbdb\" \u003e\u003cpath\u003e\u003csvg\u003e\u0026nbsp;\u0026nbsp;
未知\",
  "framework": "deepspeed",
  "framework_config": null,
  "framework_template": "",
  "ft_type": "",
  "id": 5,
  "images": "image.ccos.io/iaas_pub/notebook:jupyter-cpu",

```

```

    "lab_id": null,
    "label": "label",
    "last_run_time": "未知",
    "level": "L0",
    "lock_status": 0,
    "maas_task_type": "slm",
    "model_path": null,
    "model_type": "",
    "name": "test-train",
    "node_selector": "train=true",
    "output_dir": "",
    "percentage": "--",
    "project": 58,
    "project_id": 58,
    "resource_cpu": "2",
    "resource_gpu": "0",
    "resource_id": "AIDP-dabd0dc99a9447d3980b62c73b0078dd",
    "resource_memory": "4G",
    "retry": 0,
    "spec_type": "compute",
    "status": "",
    "storage_type": null,
    "volume_mount": "kubeflow-user-workspace(pvc):/mnt",
    "workers": 1
  },
  "status": 0
}
}

```

## 11.4 训练作业运行接口

### 11.4.1 简要描述

运行训练作业任务。

### 11.4.2 接口约束

无。

### 11.4.3 URL 格式

https://{endpoint}/api/aidp/alpha/maas\_task\_modelview/slm/api/maas\_run/{train\_work\_id}

URL 参数说明:

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/alpha	必须	string	表示接口版本号为: alpha
/slm/api/maas_run	必须	string	表示为训练作业任务
{train_work_id}	必须	Int	表示训练作业ID

### 11.4.4 请求方式

GET。

### 11.4.5 请求参数

请求 header 参数:

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型, 根据实际情况传参

### 11.4.6 返回参数

参数	参数类型	描述
message	string	接口状态信息
result	object	结果
status	string	状态码

### 11.4.7 示例

请求示例:

```
GET 'https://{endponit}/api/aidp/alpha/maas_task_modelview/slm/api/maas_run/5'  
--header 'Content-Type: application/json'
```

正常返回示例：

```
{"message": "success", "result": {}, "status": 0}
```

## 11.5 训练作业停止接口

### 11.5.1 简要描述

停止训练作业任务。

### 11.5.2 接口约束

无。

### 11.5.3 URL 格式

https://{endpoint}/api/aidp/alpha/maas\_task\_modelview/slm/api/stop/{train\_work\_id}

URL 参数说明：

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/alpha	必须	string	表示接口版本号为：alpha
/slm/api/stop	必须	string	表示为训练作业任务
{train_work_id}	必须	Int	表示训练作业ID

### 11.5.4 请求方式

GET。

### 11.5.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

## 11.5.6 返回参数

参数	参数类型	描述
message	string	接口状态信息
result	object	结果
status	string	状态码

## 11.5.7 示例

请求示例：

```
GET 'https://{endpoint}/api/aidp/alpha/maas_task_modelview/slm/api/stop/5'  
--header 'Content-Type: application/json'
```

正常返回示例：

```
{"message":"success","result":{},"status":0}
```

## 11.6 训练作业删除接口

### 11.6.1 简要描述

删除训练作业任务。

### 11.6.2 接口约束

无。

### 11.6.3 URL 格式

https://{endpoint}/api/aidp/v1/deleteInstance

URL 参数说明：

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/v1	必须	string	表示接口版本号为：v1
/deleteInstance	必须	string	表示为删除实例操作

## 11.6.4 请求方式

DELETE。

## 11.6.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

请求 param 参数：

参数	参数类型	是否必选	描述	示例
id	int	是	实例ID	5
type	int	是	删除类型	5

## 11.6.6 返回参数

参数	参数类型	描述
requestId	string	请求ID
code	string	返回码
message	string	请求结果描述
data	object	
- code	string	用户id
- message	string	用户名
- result	object	训练作业任务相关信息
- status	string	状态码

## 11.6.7 10.6.7 示例

请求示例：

```
DELETE 'https://{endponit}/api/aidp/v1/deleteInstance?id=88&type=5'  
--header 'Content-Type: application/json'
```

正常返回示例：

```
{"message": "success", "result": {}, "status": 0}
```



# 12 实验管理 API 参考

## 12.1 产品功能简介

实验管理是通过在一个实验里关联多个大模型训练任务，通过对比同一个实验里面的多个大模型训练任务的训练任务指标如：**Loss** 评估参数、**Perplexity** 评估参数来优化大模型训练过程，提高模型训练效率和效果。

## 12.2 实验管理创建接口

### 12.2.1 简要描述

创建实验管理。

### 12.2.2 接口约束

无。

### 12.2.3 URL 格式

`https://{endpoint}/api/aidp/alpha/maas_modelview/lab/api/`

URL 参数说明：

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/alpha	必须	string	表示接口版本号为：alpha
/maas_modelview/lab/api	必须	string	表示为实验管理资源

### 12.2.4 请求方式

POST。

### 12.2.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

请求 body 参数:

参数	参数类型	是否必选	描述	示例
project	int	是	项目ID	58
name	string	是	实验管理名称	lab-test
describe	string	否	实验管理描述	实验管理描述

## 12.2.6 返回参数

参数	参数类型	描述
status	int	返回状态
message	string	请求结果描述
result	object	
- id	int	实验管理id
- describe	string	实验管理描述
- name	string	实验管理名称
- project	int	实验管理项目ID

## 12.2.7 示例

请求示例:

```
POST 'https://{endpoint}/api/aidp/alpha/maas_modelview/lab/api/'
--header 'Content-Type: application/json'
--data-raw '{"project":58,"name":"lab-test","describe":"实验管理描述"}
```

正常返回示例:

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "message": "success",
  "result": {
    "describe": "\u5b9e\u9a8c\u7ba1\u7406\u63cf\u8ff0",
    "id": 2,
    "name": "lab-test",
```

```

    "project": 58
  },
  "status": 0
}

```

## 12.3 实验管理删除接口

### 12.3.1 简要描述

删除实验管理。

### 12.3.2 接口约束

无。

### 12.3.3 URL 格式

https://{endpoint}/api/aidp/alpha/maas\_modelview/lab/api/{lab\_id}

URL 参数说明：

参数	是否必选	参数类型	描述
{endpoint}	必须	string	表示请求的地址
/api	必须	string	此处为统一格式
/aidp	必须	string	表示为adip服务
/alpha	必须	string	表示接口版本号为：alpha
/maas_modelview/lab/api	必须	string	表示为实验管理资源
{lab_id}	必须	Id	表示为实验管理ID

### 12.3.4 请求方式

DELETE。

### 12.3.5 请求参数

请求 header 参数：

参数	是否必选	参数类型	示例	描述
Content-Type:	必须	string	application/json	http内容类型，根据实际情况传参

## 12.3.6 返回参数

参数	参数类型	描述
status	int	返回状态
message	string	请求结果描述
result	object	
- id	int	实验管理id
- describe	string	实验管理描述
- name	string	实验管理名称
- project	int	实验管理项目ID

## 12.3.7 示例

请求示例:

```
DELETE 'https://{endpoint}/api/aidp/alpha/maas_modelview/lab/api/2'  
--header 'Content-Type: application/json'
```

正常返回示例:

```
HTTP/1.1 200 OK  
Content-Type:application/json  
{  
  "message": "success",  
  "result": {  
    "changed_by_fk": 4,  
    "changed_on": "2025-01-21T14:03:11",  
    "created_by_fk": 4,  
    "created_on": "2025-01-21T14:03:11",  
    "describe": "\u5b9e\u9a8c\u7ba1\u7406\u63cf\u8ff0",  
    "id": 2,  
    "name": "lab-test",  
    "project_id": 58,  
    "type": "lab"  
  },  
  "status": 0  
}
```

# 13 接口集成参考

## 13.1 端到端功能简介

此篇文章以完整的端到端业务为例，将 API 串联起来，介绍了完整的开发机创建任务和大模型调优 SFT 训练任务，方便用户使用 OpenAPI 创建开发机和大模型训练。

## 13.2 Notebook创建

### 13.2.1 简要描述

端到端创建 Notebook 实例。

### 13.2.2 接口约束

无。

### 13.2.3 端到端业务流程

使用 OpenAPI 创建开发机，AK/SK 中已包含了用户信息，需要先依次：①创建项目、②购买资源、③关联计算资源、④创建开发机

#### 13.2.3.1 创建项目

参考 [9.1 项目空间创建接口](#)。

#### 13.2.3.2 购买资源

参考 [4.2 购买资源包接口](#)。

#### 13.2.3.3 关联计算资源

参考 [9.2 关联资源接口](#)。

#### 13.2.3.4 创建开发机

参考 [2.2 Notebook 创建接口](#)。

## 13.3 大模型调优SFT训练任务创建

### 13.3.1 简要描述

端到端创建大模型调优 SFT 训练任务。

### 13.3.2 接口约束

无。

### 13.3.3 端到端业务流程

使用 OpenAPI 创建大模型调优 SFT 训练任务，AK/SK 中已包含了用户信息，需要先依次：

①创建项目、②购买资源、③关联计算资源、④创建大模型 SFT 任务、⑤创建大模型调优 SFT 训练任务。

#### 13.3.3.1 创建项目

参考 [9.1 项目空间创建接口](#)。

#### 13.3.3.2 购买资源

参考 [4.2 购买资源包接口](#)。

#### 13.3.3.3 关联计算资源

参考 [9.2 关联资源接口](#)。

#### 13.3.3.4 创建大模型 SFT 任务

参考 [7.15 大模型任务创建接口](#)。

#### 13.3.3.5 创建大模型调优 SFT 训练任务

参考 [7.2 大模型预训练、调优\(SFT、DPO、RLHF\)创建接口](#)。