

腾讯云 TI 平台 TI-EMS

操作指南

产品文档



腾讯云

【 版权声明 】

©2013–2022 腾讯云版权所有

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

【 商标声明 】



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

【 服务声明 】

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

【 联系我们 】

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

文档目录

操作指南

控制台说明

概览

模型仓库

创建模型服务配置

管理模型服务配置

模型运行环境

自定义运行环境

准备自定义运行环境

导入自定义运行环境

模型优化

模型服务

在线推理

批处理作业

服务调用方式

监控与日志

监控与告警

日志分析

资源组管理

操作指南

控制台说明

最近更新时间：2021-12-22 11:41:08

控制台概述

使用 TI-EMS，您需要先登录 [腾讯云 TI 平台 TI-EMS 控制台](#)，控制台页面左侧栏包括**概览**、**模型仓库**、**模型服务和资源组管理**四部分。

概览

概览页展示了系统资源的使用情况和服务调用情况。

模型仓库

模型仓库包含**模型服务配置**和**模型优化**两个模块，**模型服务配置**管理了不同版本的模型，用户可以在此页面创建一个模型服务配置，选择模型服务启动的地域，定义模型文件和服务运行环境，然后在列表页面可将该模型服务配置启动为在线推理服务或批处理作业。**模型优化**模块提供了模型优化功能，即把已有的模型转换成所需格式，以便应用于算力和性能更高的资源上。

模型服务

TI-EMS 提供在线推理和批处理作业两种模型服务，可在模型服务配置列表启动相应的服务类型，启动完成后可在**模型服务>在线推理**或**模型服务>批处理作业**页面查看服务运行情况。在启动在线服务的过程中，可以根据业务需求采取手动或自动的方式调节实例伸缩范围，从而达到对在线服务的精细化控制与管理。

资源组管理

在 TI-EMS 中，集群资源会被分为不同的资源组进行隔离，在启动模型服务的时候，用户可以选择部署在默认的公共资源组或者部署在用户独享的专用资源组中。部署在不同资源组的模型服务计费方式详见 [计费概述](#)。

资源组管理还提供手动或自动扩展节点的功能，资源组详细功能请见 [资源组管理](#)。

概览

最近更新时间：2020-06-29 17:02:33

概览页展示了系统主要资源使用情况和运行服务情况。

主要 KPI 指标解释：

目前 TI-EMS 仅支持北京、广州、上海三个地域，切换顶部的地域栏，可查看不同 region 下的统计数据。

资源概况

- **公共资源组-CPU 使用：**当前用户公共资源组运行的服务占用的总 CPU 核数。
- **公共资源组-GPU 使用：**当前用户公共资源组运行的服务占用的总 GPU 卡数。
- **公共资源组-MEM 使用：**当前用户公共资源组运行的服务占用的总内存量。
- **专用资源组-CPU 使用：**当前用户专用资源组运行的服务占用的总 CPU 核数。
- **专用资源组-GPU 使用：**当前用户专用资源组运行的服务占用的总 GPU 卡数。
- **专用资源组-MEM 使用：**当前用户专用资源组运行的服务占用的总内存量。

服务概况

- **总服务配置数量：**统计了当前用户所有服务配置的数量（包括所有配置的不同版本）。
- **总运行服务数量：**统计了当前用户的总运行中的服务数量（包括在线推理和批处理作业）。
- **运行中/总在线推理数量：**统计了当前用户运行中/所有状态的在线服务数量。
- **运行中/总批处理作业数量：**统计了当前用户运行中/所有状态的批处理作业数量。
- **Top10服务调用量：**统计了当前用户所有运行服务中请求数量为前10的服务，单击任一服务，可查看其对应的服务请求曲线，同时也可以查看服务对应的配置情况和资源使用情况。

🔍 说明：

服务运行监控 Dashboard 可切换时间粒度分别为近3个小时、昨天、近3天、近5天、近7天和自定义时间区间。

模型仓库

创建模型服务配置

最近更新时间：2022-01-10 17:16:37

进入模型仓库>模型服务配置，单击新建，开始创建模型服务配置。

模型服务配置

北京

广州

上海

如果遇到产品相关问题，您可 [提交工单](#) 寻求帮助。欢迎加入腾讯云TI-EMS用户群进行交流讨论，群号：836843676。

新建

auto_conf_angel
2019-12-13 10:14:00

1.0

在线推理

批处理作业

更多

运行环境 angel
模型文件 cos://ti-ems-1255502019.cos.ap-beijing.myqcl...
版本说明 auto_config_angel

demo_openvino
2019-12-12 19:33:18

1.0

在线推理

批处理作业

更多

运行环境 openvino
模型文件 cos://ti-ems-1255502019.cos.ap-beijing.myqcl...
版本说明

创建 TI-EMS 模型服务配置，需要：

1. 输入基本配置信息（模型服务配置的名称、地域、模型服务的运行环境）
2. 配置模型文件

基本配置信息

名称：要创建的模型服务配置的名称，不超过20个字符。

地域：选择服务所运行的地域。

版本：新建模型服务配置版本自动生成，为1.0，后续每新增一个版本，版本号自动加1。

运行环境：单击运行环境，可选择模型运行的公共环境。TI-EMS 提供 TF-Serving、PMML、TensorRT、

OpenVINO 和 Angel 五种模型服务运行环境，相关介绍请详见 [模型运行环境](#)。

运行环境



公共运行环境

我的运行环境

<input type="radio"/> pmml
<input type="radio"/> angel
<input type="radio"/> tfserving
<input type="radio"/> tensorrt
<input type="radio"/> openvino

确定

取消

模型文件配置

TI-EMS 支持三种模型文件输入方式：

1. 直接输入 COS 路径
2. 从 COS 存储桶选择模型文件
3. 本地上传模型文件

直接输入 COS 路径

您可以直接在模型文件输入框输入模型文件夹 COS 路径，其中模型地址格式为：

`cos://{bucket_name}-${appid}.cos.{region}.myqcloud.com/模型文件夹路径/`

示例

假设用户通过主账号（AppID 为1255502019）登录 COS 控制台进行创建存储桶，存储桶的配置信息为：地域选择为北京地域，存储桶名称自定义为 ti-ems，模型文件夹为 inception（[单击下载模型文件夹](#)，上传至存储桶 ti-ems 的 models/tfserving 路径下，如下图所示：

[< 返回](#) | ti-ems

文件列表

基础配置

域名管理

文件碎片列表

+ 上传文件

创建文件夹

批量删除

[< 返回上一级](#) | [models/tfserving/inception](#)

<input type="checkbox"/>	文件名	大小
<input type="checkbox"/>	 variables	--
<input type="checkbox"/>	saved_model.pb	806.80KB

则存储桶默认域名如下所示：

```
cos://ti-ems-1255502019.cos.ap-beijing.myqcloud.com/models/tfserving/inception/
```

其中：

- ti-ems-1255502019：表示该存储桶归属于 AppID 为1255502019的用户。AppID 是您在成功申请腾讯云账户后所得到的账号，由系统自动分配，具有固定性和唯一性，可在 [账号信息](#) 中查看。
- cos：指对象存储 COS。
- ap-beijing：指存储桶的地域简称。
- myqcloud.com：腾讯云域名，固定字符。
- models/tfserving/inception/：模型文件夹路径。

从 COS 存储桶选择

TI-EMS 控制台可直接选择 COS 模型文件夹，在模型服务配置新建页面，单击**对象存储 COS 文件**，进入 COS 文件选择弹窗页面。

名称	<input type="text" value="请输入的名称, 1~20字符"/>
地域	<input type="button" value="北京"/> <input type="button" value="广州"/> <input type="button" value="上海"/> 仅支持您当前选择地域的COS存储桶
运行环境	<input type="button" value="请选择运行环境"/> 运行环境
模型文件	<input type="text" value="可以选择对象存储COS中的文件, 或选择本地上传"/> 对象存储cos文件
	前往 cos控制台 上传文件, 关于更多文件路径规则说明, 请查看 产品文档
配置版本	<input type="text" value="1.0"/>

选中模型文件夹所在的路径, 单击**确定**, 完成模型文件配置。

对象存储cos文件 ✕

◀ ▶

✕ 🔍
☰

文件名
📁 variables
📄 saved_model.pb

本地上传模型文件

TI-EMS 还可支持在控制台直接上传本地模型文件夹, 无须跳转至对象存储页面进行操作。在模型服务配置新建页面, 点击**对象存储 COS 文件**, 在 COS 文件选择弹窗页面, 进入对应的目录, 单击**上传**, 即可将本地模型文件夹上传至 COS 存储桶中, 上传完成后, 选中模型文件夹路径, 单击**确定**, 完成模型文件配置。

⚠ 注意:

1. 模型服务部署所在的地域与模型文件所在 COS 存储桶地域必须一致。例如, 您需要将服务部署在北京地域, 请将模型文件上传至北京地域存储桶。
2. TI-EMS 所约定的模型文件路径, 是模型文件所在的文件夹的路径。
3. TI-EMS 目前支持 TensorFlow SavedModel、PMML、Angel、OpenVINO IR 四种格式的模型文件。
4. 对象存储 COS 相关操作请详见 [COS 产品文档](#)。

管理模型服务配置

最近更新时间：2021-12-22 11:46:20

查看模型服务配置

模型服务配置基本信息：

在模型服务配置列表页面，您可以查看不同模型服务配置版本的基本信息，包括运行环境信息、模型文件、版本说明等。

相关操作：

单击**在线推理**或**批处理作业**，您可以对模型服务配置进行启动在线推理或批处理作业服务操作，单击**更多>新增版本**，您可以新增模型服务配置版本，单击**更多>删除**，您可以选择删除模型服务配置。

新增配置版本

单击**新增版本**，您可以新增模型服务配置版本，新增的配置版本号自动加1，新增版本不能修改原配置名称，您可以在新的模型服务配置版本中更新您的模型文件，更方便地进行模型管理。

模型运行环境

最近更新时间：2021-12-24 15:55:10

目前，TI-EMS 提供五种模型运行环境：

- TF-Serving：Tensorflow v1.13.1，业界流行的 TensorFlow Serving 模型服务，支持 SavedModel 模型。
- PMML：通用标准的 PMML 模型服务，支持标准 PMML 模型。
- TensorRT：Tensorflow v1.13.1，支持模型自动优化的 Tensorflow 模型服务镜像，支持 SavedModel 模型。
- OpenVINO：OpenVINO-model-server v2019.1.1，专门针对 CPU 优化的模型服务镜像，支持 OpenVINO Intermediate Representation(IR) 模型。
- Angel：腾讯自研的 Angel 框架，支持 PMML 和 Angel 模型。

TF-Serving

TI-EMS 提供的 TF-Serving 推理运行环境可以加载 Tensorflow SavedModel 模型。

PMML

PMML (Predictive Model Markup Language) ，一种预测模型标记语言，TI-EMS 可以加载标准的 PMML 模型。

TensorRT

TensorRT 镜像是支持模型自动优化的 Tensorflow 模型服务镜像,支持用户使用 gRPC 和 HTTP 访问模型服务，访问方式兼容 TF-Serving 镜像。TensorRT 支持用户上传 Tensorflow SavedModel 模型，系统自动对模型进行优化。对8种主流分类模型测试，TensorRT 镜像推理性能相对 TF-Serving 镜像平均提升60%以上。对5种主流目标识别模型测试，平均性能提升40%以上。

上述测试模型详情如下：

模型类型	模型名称
分类	Inception_v3、Inception_V4、MobileNet_V1、NasNet_Mobile、NasNet_Large、ResNet_V2_50、Xception、DenseNet169
目标检测	IFaster_R-CNN_Nas、Mask_R-CNN_ResNet_50、SSD_Inception_V2、SSD_MobileNet_V1、SSD_ResNet_50_FPN

使用 TensorRT 镜像可能遇到问题说明：

- 模型启动时间较长：因为需要做模型优化，模型启动时间可能会比较长。对于某些大模型，例如 NasNet，模型启动时间可能需要3到5分钟。

- 模型首次推理时间较长：系统会根据用户上传数据大小生成最优代码，模型首次推理时间（仅限首次推理时间）可能需要0.5 – 5分钟。
- 模型加载失败：因为 Tensorflow 对读取 protobuf 文件大小限制，单个 protobuf 文件大于1GB会读取失败。即使上传模型小于1GB，模型优化过程也可能会生成大于1GB模型文件而导致加载失败。
- 模型服务失败：Tensorflow 自动模型优化功能在快速迭代中，某些优化后的 SavedModel 可能会服务失败，请通过 [腾讯云 TI 平台 AI 开发者社区](#) 讨论相关问题，腾讯云 AI 专家会及时帮助解答。
- TensorRT 镜像相对 TF-Serving 镜像推理性能提升不明显。请联系腾讯云客服，针对大客户，腾讯云 AI 专家可以针对具体模型深度优化。

OpenVINO

OpenVINO 镜像是专门针对 CPU 优化的模型服务镜像，支持用户上传 OpenVINO Intermediate Representation (IR) 模型，支持 HTTP 和 gRPC 访问模型服务，模型访问方式兼容 SavedModel 镜像。对5种主流分类模型测试，OpenVINO 镜像推理性能相对 TF-Serving 镜像平均提升2.4倍以上。对4种主流目标识别模型测试，OpenVINO 镜像推理性能相对 TF-Serving 镜像平均性能提升30%以上。

上述测试模型详情如下：

模型类型	模型名称
分类	Inception_v1、Inception_v3、VGG-19、ResNet_V1_50、NasNet_Large
目标检测	Mask_R-CNN_Inception_V2_COCO、RFCN_ResNet_101_COCO、MobileNet-SSD_V1_COCO、Faster_R-CNN_Inception_V2_COCO

使用 OpenVINO 镜像可能遇到问题说明：

- 模型推理结果有误：因为 OpenVINO 模型转换包括数据预处理，请检查模型转换过程参数设置，本地验证 IR 推理结果正确后再上传模型。
- OpenVINO 镜像推理性能问题。请联系腾讯云客服，针对大客户，腾讯云 AI 专家可以针对具体模型深度优化。
- 如何将 Tensorflow/Caffe/MXNET/ONNX 等模型转换为 OpenVINO IR 模型，请参考：[模型格式转换](#)。
- HTTP 接口调用仅支持数值数据类型，详情请参考：[OpenVINO 接口调用说明](#)。

Angel

腾讯自研深度学习框架，Angel 相关内容请访问 [腾讯自研的 Angel 框架](#)。

自定义运行环境

准备自定义运行环境

最近更新时间：2021-12-22 11:46:38

TI-EMS 可支持用户使用自定义运行环境，您可以在 TI-EMS 上导入自定义镜像，并且使用您的镜像加载模型。单击**运行环境**，您将看到**我的运行环境**页面，在这个页面您可以添加自定义镜像，并加载模型生成服务。

TI-EMS 自定义运行环境规范

- **Docker 镜像**：需用户制作 Docker 镜像，解决运行服务脚本需要的依赖和环境变量，平台获取镜像权限后执行 `docker run [docker-image] /usr/bin/start.sh`，用户制作镜像时需要提供 `/usr/bin/start.sh`。
- **环境变量来源**：用户制作镜像时设置（平台不负责设置）。
- **模型文件**：用户提供模型地址，映射为 `/data/model/m/`。
- **服务脚本**：需用户开发，HTTP 端口9001，gRPC 端口9000。
- **处理请求**：需用户开发，处理端口9001上的 HTTP POST 请求，处理端口9000上的 gRPC 请求。
- **健康检查**：需用户开发，处理端口9001上的 HTTP GET 请求，并返回 AVAILABLE。
- 需要用户开启鉴权

创建步骤

首先创建 Docker 基础镜像，包括安装相关框架以及依赖包。然后迁移模型服务代码到基础镜像，需根据 TI-EMS 模型文件路径以及模型服务端口约定修改模型服务代码。最后，添加服务启动脚本，TI-EMS 启动会默认运行模型启动脚本。

1. 创建 Docker 基础镜像，编写 **Dockerfile** 文件。
2. 迁移模型服务脚本，根据 TI-EMS 模型文件以及模型服务端口约定修改模型服务脚本。
 - 2.1 修改模型文件加载路径，以及模型服务脚本中和模型文件路径相关代码，TI-EMS 模型文件默认保存路径为：`/data/model/m`。
 - 2.2 修改模型服务端口，TI-EMS HTTP 模型服务使用9001端口，gRPC 服务默认是用9000端口。
 - 2.3 添加启动脚本，TI-EMS 启动容器会默认调用 `/usr/bin/start.sh`。
3. 将修改后模型服务文件，和启动脚本文件放到基础镜像对应位置，`commit` 生成最终 TI-EMS 服务镜像（迁移模型服务脚本可能需要多次修改，建议先生成基础镜像，然后修改模型服务脚本和添加启动脚本等，调试完成后，再 `commit` 生成最终镜像）。

有更多问题，请加入 TI-EMS 用户交流群（QQ 群：836843676）或直接点击 [在线咨询](#)。

导入自定义运行环境

最近更新时间：2021-12-22 11:46:43

将自定义镜像推送到镜像仓库

TI-EMS 支持加载腾讯云容器服务镜像仓库中的镜像，在使用自定义运行环境功能之前，您需要将您的镜像 docker push 到腾讯云容器服务镜像仓库。

⚠ 注意：

目前 TI-EMS 仅能拉取镜像源为 public 的镜像，若出于各方面原因，您无法将镜像设置为 public，请[联系我们](#)。

1. 进入腾讯云 [容器服务控制台](#)。
2. 进入[镜像仓库](#)，创建命名空间和您的个人镜像仓库，具体操作方法详见 [镜像仓库基本教程](#)。
3. 登录腾讯云 docker registry。

```
sudo docker login --username=UIN ccr.ccs.tencentyun.com
```

- UIN：您的账号 ID。
 - 密码：您开通 namespace 时设置的密码。
4. 将镜像推送到 registry。

```
sudo docker tag [ImageId] ccr.ccs.tencentyun.com/[namespace]/[ImageName]:[镜像 版本号]
sudo docker push ccr.ccs.tencentyun.com/[namespace]/[ImageName]:[镜像版本号]
```

- ImageId 和镜像版本号根据镜像信息补充。
- namespace 是开通镜像仓库时填写的命名空间。
- ImageName 是在控制台创建的镜像名。

在 TI-EMS 中导入自定义镜像

1. 在 TI-EMS 自定义运行环境页面单击新增环境，输入：
 - 运行环境名称。
 - 运行环境地址，格式如下：

```
ccr.ccs.tencentyun.com/[namespace]/[ImageName]:[镜像版本号]
```

-
- 描述信息
2. 保存信息后，选中对应的自定义环境，单击**确定**。
 3. 选择对应的模型文件，模型文件路径详见 [创建模型服务配置](#)。

模型优化

最近更新时间：2022-01-10 17:24:32

TI-EMS 提供了模型量化功能，针对 Tensorflow SavedModel 格式的模型文件，希望在英伟达 GPU 上获得更高的推理性能。可以使用模型优化操作，对您的模型进行优化。优化后的模型可以部署在带有 Nvidia P4 或 T4 卡的机器上。

目前模型优化使用基于 TensorRT 的模型转换。包括 FP32 模型到 FP32 模型转换，FP32 模型到 FP16 模型的转换和 FP32 模型到 INT8 模型的转换。

创建模型量化任务

1. 登录 [TI-EMS 控制台](#)，在左侧导航栏中选择 **模型仓库 > 模型优化**，进入模型优化页面。
2. 单击左上角“创建量化任务”，进入任务创建页面。

模型优化 北京 产品文档 API文档

如果遇到产品相关问题，您可[提交工单](#)寻求帮助。欢迎加入腾讯云TI-EMS用户群进行交流讨论，群号：836843676。

创建量化任务

任务名称	任务状态	转换模板	描述	创建时间	操作
mtt_test	运行成功	Tensorflow saved_model转换成Tensorrt		2019-12-17 15:55:26	删除
mt_test	等待中	Tensorflow saved_model转换成Tensorrt	测试一下模型量化的功...	2019-12-17 15:43:46	删除

共 2 项 每页显示行 10 1 / 1 页

3. 在创建任务页面，输入以下信息：
 - 任务名称：1-20字符。
 - 任务描述：输入您对该次量化任务的备忘信息。
 - 选择资源组：目前模型量化任务仅支持部署在专用资源组，下拉选择合适的专用资源组。
 - 实例配置：TensorRT 量化任务一般需要 GPU 资源，请选择**GPU 配置**，并选择内存大于8G的资源。
 - 转换模板：TI-EMS 目前仅支持将 Tensorflow SavedModel 转换成基于 TensorRT 的模型，使其可在 GPU 上进行推理。
 - 转换输入目录：请查看[模型输入格式说明](#)。
 - 转换输出目录：请选择量化后模型需要输出的目录。
 - 量化精度：选择量化精度。
 - 量化批大小：选择量化批大小。
4. 任务创建完成后，单击**启动任务**，完成量化任务的创建。单击量化任务的名称，可进入任务详情页，切换至**日志**选项卡，可查看任务运行过程中的日志。量化任务可以进行删除。

模型输入格式说明

```
|  
|---model 模型存放目录，必须以model命名。  
  
|---saved_model.pb 模型文件。  
|---variables 变量存储文件夹。  
  
|---data 转INT8模型所需校准数据集存放目录，必须以 data 命名。FP16和FP32转换不需要。输入目录下有  
且只能有一个。  
  
|---data.npy 校准数据集为npy格式文件，文件名为data.npy。npy是在预处理后直接输入模型的数据。假设  
模型输入图片大小为{1, 3, 299, 299},数据集大小为100张图片。data.npy的格式为{100, 3, 299, 299  
}。
```

很少情况下 Tensorflow SavedModel 可能包含多个模型，模型优化会默认优化标记为“serve”的模型。目前模型优化仅支持单输入模型。

如果模型优化操作成功结束，模型优化后输出为单个 pb 文件。优化后的模型可以和未优化模型一样使用 Tfserving 环境部署。

鉴于目前 TensorRT 支持的模型有限，如果您的模型不能成功使用模型优化功能，并且希望提升模型推理性能，您可单击 [在线咨询](#) 寻求帮助。

模型服务 在线推理

最近更新时间：2021-12-22 11:47:05

配置服务启动基本信息

服务名称：创建服务的名称。

资源组：选择服务运行占用的资源组（资源组详细介绍请看[资源组管理](#)）。

服务配置：本服务对应的模型服务配置信息，由[创建模型服务配置](#)步骤确定。

实例资源配置

TI-EMS 针对不同的模型服务运行资源需求为您提供了 **CPU 配置**和 **GPU 配置**，若您选择在公共资源组启动服务，则需要选择 CPU quota 和 GPU quota 并指定对应 quota 的数量，公共资源组的服务计费方式按照 quota 单价 * 数量 * 服务运行时长；若您选择在专用资源组启动服务，则需自定义 CPU 配置和 GPU 配置，用户在预购的专用资源组启动的服务不再另外收费。公共资源组和专用资源组的收费情况请详见[计费概述](#)。

配置实例扩展策略

手动调节实例：选择手动调节实例，可以直接设定启动服务时的实例数量，实例数量最小设置为1。

自动调节实例：选择自动调节实例，可以设置多种触发策略，当满足触发策略时，按照指定的实例伸缩范围自动调节。TI-EMS 目前支持的触发指标如下：

- CPU 利用率
- 内存利用率
- GPU 利用率

当系统实际指标小于触发指标设定值时，执行自动缩容，大于设定值时，执行自动扩容。实例配置完成后，单击**启动服务**，即可快速启动服务，并跳转到模型服务列表页面。

更新在线服务

单击模型服务列表页面的**更新**，即可更新当前服务的配置，单击服务配置下拉框，您可以选择该服务配置的其他不同版本，您也可以改变服务启动的实例调节触发策略。

服务配置更新完成后，单击**启动服务**，即可快速更新一个服务，并跳转到模型服务列表页面。

管理在线服务

停止与启动服务

单击列表右侧的**停止**，正在运行的服务会启动停止程序，服务所占用的资源即被释放。单击**启动**，服务会按照当前配置项进行重新启动。

删除服务

单击列表右侧的**更多**>**删除**，即可删除对应的模型服务。

🔗 说明：

正在运行的服务不可以直接删除，需要先停止该服务，才可以删除。

批处理作业

最近更新时间：2021-12-22 11:47:13

启动批处理作业

前提条件

- 准备好需要批量处理的数据，并上传至 COS 目录。
- 已在 COS 创建至少1个空文件夹，用户存储批处理作业预测输出的内容。

配置服务启动基本信息

服务名称： 创建服务的名称。

资源组： 选择服务运行占用的资源组（资源组详细介绍请看 [资源组管理](#)）。

模型服务配置： 本服务对应的模型服务配置信息，由 [创建模型服务配置](#) 步骤确定。

输入数据： 选择输入数据的 COS 路径，即您上传数据的 COS 目录，只能选择文件夹。

数据类型： 用户输入数据的输入格式类型，给定作业的所有输入数据必须使用相同的数据格式，目前仅支持 JSON 格式的数据。

Batchsize： 选择用户输入数据的批处理大小。

输出数据： 选择批量预测结果的保存位置，可以选择您创建的空文件夹。

实例资源配置

TI-EMS 针对不同的模型服务运行资源需求为您提供了 **CPU 配置**和 **GPU 配置**，若您选择在公共资源组启动服务，则需要选择 CPU quota 和 GPU quota 并指定对应 quota 的数量，公共资源组的服务计费方式按照 $quota \text{ 单价} * \text{数量} * \text{服务运行时长}$ ；若您选择在专用资源组启动服务，则需自定义 CPU 配置和 GPU 配置，用户在预购的专用资源组启动的服务不再另外收费。

设定实例数量： 设定批处理作业启动时占用的实例数量，实例数量最小设置为1。

实例配置完成后，单击**启动服务**，即可快速启动批处理作业，并跳转到批处理作业列表页面。

管理批处理作业

停止作业

单击列表右侧的**停止**，正在运行的作业会启动停止程序，服务所占用的资源即被释放。

注意：

已停止的批处理作业不可再次启动。

删除作业

单击列表右侧的**更多>删除**，即可删除对应的批处理作业。

注意:

正在运行的作业不可以直接删除，需要先停止该作业，才可以删除。

日志

TI-EMS 支持查看单个作业不同实例的日志，用户可通过日志数据窗口查看各作业的运行日志。

服务调用方式

最近更新时间：2021-12-08 18:06:47

公网地址调用

TI-EMS 的在线服务一旦部署完成后，您可以选择创建一个公网地址来进行服务调用。调用方式为 gRPC 或 HTTP，调用地址和需要的鉴权信息可在控制台对应的在线服务列表处单击[更多>调用>公网调用查看](#)。

服务调用 demo 详见 [最佳实践](#)。

VPC 地址调用

TI-EMS 支持用户自己的 VPC 通过 privatelink 通道和 TI-EMS 的 VPC 环境打通。在控制台对应的在线服务列表处单击[更多>调用>VPC 调用](#)，选择需要打通的 VPC 私有网络和子网信息，完成打通后可显示调用地址和需要的鉴权信息。VPC 相关操作请详见 [VPC 产品文档](#)。

VPC 高速直连调用

除了通过 privatelink 打通两个 VPC，TI-EMS 还可通过挂载用户弹性网卡到服务容器（专用资源组）上，来实现两个 VPC 中网络环境的互通。下面介绍 VPC 高速直连的调用方式。

准备环境

腾讯云官网提供了各种语言的 SDK 供使用，这里我们使用 Python SDK 做演示。

您可以通过 pip 安装 Python SDK。

```
pip install tencentcloud-sdk-python
```

使用 sdk 之前您需要在控制台获取 SecretID、SecretKey。

调用示例

```
from tencentcloud.common import credential
from tencentcloud.common.exception.tencent_cloud_sdk_exception import TencentCloudSDKException
from tencentcloud.tiems.v20190416 import tiems_client, models
import requests

def do_request(address, request_str):
    headers = {"Content-type": "application/json", "Accept": "application/json", "X-AUTH-TOKEN": ""}
    try:
```

```
response = requests.post(address, data=request_str, headers=headers)
print(response.text)
except Exception as e:
    print(e)
if __name__ == '__main__':
    try:
        # 实例化一个认证对象，入参需要传入腾讯云账户secretId, secretKey
        cred = credential.Credential("xx", "yy")
        # 实例化client对象
        client = tiems_client.TiemsClient(cred, "ap-beijing")
        # 实例化一个请求对象 并请求查询服务，这里我们提前在页面上创建了一个 两副本的 pmml 服务，并且做了内网打通
        req = models.DescribeServicesRequest()
        resp = client.DescribeServices(req)
        # 输出json格式的字符串回包
        replicas = resp.Services[0].Status.ReplicaInfos
        ips = [a.EniIp for a in replicas if a.Status == "Normal"]
        # 发送预测请求，这里 我们做了一个简单的负载均衡，依次访问两个副本ip
        print(ips)
        for ip in ips:
            address = 'http://' + ip + '/v1/models/m:predict'
            do_request(address, '{"instances" : [{"x1": 6.2, "x2": 2.2, "x3": 1.1, "x4": 1.}]}')
        except TencentCloudSDKException as err:
            print(err)
```

```
['10.20.142.88', '10.20.142.79']
{"predictions": [{"probability(1)": "0.07345460015950317", "probability(-1)": "0.9265453998404969", "y": "-1"}]}
{"predictions": [{"probability(1)": "0.07345460015950317", "probability(-1)": "0.9265453998404969", "y": "-1"}]}
```

⚠ 注意:

TI-EMS 目前仅支持专用资源组 VPC 打通。

API Gateway 公网调用

TI-EMS 新增 API Gateway 公网地址调用功能，该功能需要您开通 API 网关服务，使用 API Gateway 公网地址调用功能，系统会在模型服务部署完成后自动帮您生成一个 API 网关层面的公网地址，您可以作为服务请求地址调用，该功能正在内测阶段，请 [单击申请](#) 进行体验。

监控与日志

监控与告警

最近更新时间：2022-01-10 17:25:20

监控指标说明

TI-EMS 提供以下监控指标，您可以在 TI-EMS 控制台或者腾讯云云监控控制台进行监控信息查看。

指标名称	指标含义	单位	维度
QPS	每秒响应的请求数量，按所选的时间粒度统计求和	次	服务/ 实例
异常请求	指一段时间的异常请求数量，这里的错误指的是系统内部错误（HTTP 5xx），按所选的时间粒度统计求和	次	服务/ 实例
平均响应时间	处理请求的平均响应时间，按所选的时间粒度取平均值	秒	服务/ 实例
网络流量	每秒字节数	Bps	服务/ 实例
CPU 使用率	服务运行所占用的 CPU 资源	%	服务/ 实例
内存使用率	服务运行所占用的内存资源	%	服务/ 实例
GPU 使用率	服务运行所占用的 GPU 资源	%	服务/ 实例
显存使用率	服务运行所占用的显存资源	%	服务/ 实例

监控数据

方式1：在 TI-EMS 控制台查看监控数据

首先进入[在线服务列表](#)页面，单击[服务名称](#)，进入服务二级详情信息页，在[监控信息](#)页面，可以查看该服务的监控指标数据，TI-EMS 控制台支持服务和实例维度的监控信息查看。

The screenshot shows the monitoring page for service 'hxtestmonitor1'. It includes a navigation bar with tabs for '基本信息', '实例列表', '监控信息', '事件信息', and '运行日志'. A blue banner at the top provides a tip: '如果当前服务尚未配置告警服务, 可以前往新增告警 为服务添加告警策略'. Below this, there are time range filters ('近1小时', '近24小时', '近7天', '选择日期') and a '时间粒度: 1分钟' dropdown. A dropdown menu is open, showing the selected service 'hxtestmonitor1' and two instance IDs: 'azkwhpmwfj57vg5q-6f6b4bf7f7-82knz' and 'azkwhpmwfj57vg5q-b96c49b4f-bcrqc'. A table displays metrics: '每秒响应请求次数 (QPS)' with values 1, 0.8, 0.6, 0.4, 0.2, and 0. A summary table shows 'Max: 0.0', 'Min: 0.0', and 'Avg: 0.0'. A line chart at the bottom shows a flat line at 0.

方式二：在云监控控制台查看监控数据

您可以登录 [云监控](#) 查看 TI-EMS 在线服务监控信息。

The screenshot shows the '智能弹性模型服务' (Intelligent Elastic Model Services) page in the Cloud Monitoring console, filtered by '广州' (Guangzhou). A search bar at the top right contains the text '多个关键字用竖线"|"分隔, 多个过滤标签用回车键分隔'. Below the search bar is a table with the following data:

服务名称	监控	区域	创建时间
hxmonitor1 a5mf2vdm2vpvwm6	📊	ap-guangzhou	2020-03-26T17:33:55+08:00
auto_angel_tecpu-MANUAL-HT ax9nt64zdztsqj4	📊	ap-guangzhou	2020-03-26T10:04:53+08:00
test arfr4z4fh2qkg9v2	📊	ap-guangzhou	2020-03-25T11:36:11+08:00

配置告警

您可以通过 [云监控](#) 对模型服务配置告警策略，从而对服务运行状态进行监控，目前 TI-EMS 可以配置的告警指标包括上述所有监控指标，指标说明请查看 [监控指标说明](#)。同时，告警支持选择接收告警的用户组，选择接收邮件、短信、微信等方式的告警渠道。

1. 登录 [云监控](#) 控制台，在左侧导航栏中，选择[告警配置](#)>[告警策略](#)，单击[新建](#)，开始配置告警策略。
2. 在[新建策略](#)页面中，根据实际需求，填写策略名称，选择策略类型，以及告警对象等。
 - 策略名称：自定义

- 策略类型：选择“腾讯云 TI 平台 TI-EMS”
 - 告警对象：根据实际需求进行设置。若选择“选择部分对象”，可以选择不同地域下的模型服务，请选择您需要适用告警策略的模型服务。
3. 根据实际需求，设置“触发条件”和“告警策略”，如下图所示：
4. 单击**完成**，即可在**告警策略**中，查看到已配置好的策略，并可根据实际需求，选择随时启停。更多关于告警策略配置的说明，请参考 [告警策略](#)。

日志分析

最近更新时间：2021-12-22 11:48:12

在 TI-EMS 模型服务运行过程中，会产生大量日志，TI-EMS 提供本地日志查看和投递至腾讯云日志服务两种日志分析场景功能。

本地日志查看

1. 进入 [在线推理](#) 或 [批处理作业](#) 页面，单击**服务名称**，进入相关模型服务二级详情页。
2. 在相关模型服务二级详情页单击**运行日志**，进入日志查看页面。您可以在此查看近24小时的日志数据。除此之外，您还可以通过切换左侧实例 ID 查看该服务下不同实例的日志。

← pmml1234

基本信息 实例列表 监控信息 事件信息 **运行日志**

ac4s2j7w56kf2h7q-74877b8458-gdwmw ▾ 实时 近24小时 重置 前往日志分析服务 [🔗](#)

```
1 + checkVariable SECRET_ID, AKIDQ5gpark7eozE7dgAZhVS-cwZh9DdMntLmK-7rcF0MQ6PyqQwNTekB-50wNg-hPWq
2 + '[' -z AKIDQ5gpark7eozE7dgAZhVS-cwZh9DdMntLmK-7rcF0MQ6PyqQwNTekB-50wNg-hPWq ]]
3 + checkVariable SECRET_KEY, 'i04SwxubRE4rzjK8pIjHKCVxPCqbqhwoY5KXrk1r9Q='
4 + '[' -z 'i04SwxubRE4rzjK8pIjHKCVxPCqbqhwoY5KXrk1r9Q=' ]]
5 + checkVariable BUCKET, ti-ems-1255502019
6 + '[' -z ti-ems-1255502019 ]]
7 + checkVariable ENDPOINT, cos.ap-beijing.myqcloud.com
8 + '[' -z cos.ap-beijing.myqcloud.com ]]
9 + checkVariable REMOTE_PATH, models/pmml/
10 + '[' -z models/pmml/ ]]
11 + checkVariable LOCAL_PATH, /data/model/m
```

将日志投递至 CLS

开通日志服务

目前日志服务处于公测阶段，使用服务需要提前开通白名单。进入 [腾讯云日志服务 CLS](#) 产品介绍页，单击**立即申请**开通白名单，业务描述请填写"TI-EMS 日志分析检索"（一般审核时间在1-2个工作日，请耐心等待）。

若您已之前已开通日志服务白名单，即可跳过此步骤。

开启日志投递

1. 角色授权

启动模型服务时，您可以选择将服务日志投递至 CLS，单击日志投递启用按钮。首次使用日志服务投递功能需要进行角色策略授权，单击**确定授权**，跳转至服务授权页面，单击**同意授权**，即完成日志投递授权，同时页面自动跳转至 TI-EMS 控制台。

注意：

若当前用户为子账号或者协作者，需要联系主账号或者拥有 QcloudCamFullAccess 权限的子账号为您当前的子账号/协作者用户授予 QcloudCamSubaccountsAuthorizeRoleFullAccess 策略，使您的子账号/协作者账号拥有服务角色授权权限。

← 角色管理

服务授权

同意赋予 **智能钛弹性模型服务** 权限后，将创建服务预设角色并授予 **智能钛弹性模型服务** 相关权限

角色名称 TI-EMS_QCSRole

授权策略名 QcloudAccessForTIEMSRoleInLogSearchAndQuery

角色类型 服务角色

角色描述 智能钛弹性模型服务(TI-EMS)操作权限含列举对象存储(COS)文件、含增删查改COS文件内容；含列举私有网络(VPC)、子网；含API网关全读写权限等；含增删改查日志服务(CLS)日志主题、日志集等。

同意授权

取消

2. 创建/选择日志集和日志主题

日志投递需要指定投递的日志集和日志主题，您可以选择**一键创建默认投递日志集 TI-EMS 和默认日志主题 Services**，或者您可以下拉选择将日志投递至您个人创建的日志集和日志主题中。

▲ 高级配置

实例调节 · 手动调节 直接设定实例数量

自动调节 系统实际指标小于触发指标设定值时，执行自动扩容；大于设定值时，自动扩容

实例数量 · 1 个

日志投递 ⓘ 启用后可通过一键创建，生成默认日志集/日志主题用于将服务日志将投递至 [腾讯云CLS日志服务](#)

CLS日志 ⓘ

使用默认日志集(TI-EMS)

使用默认日志主题(Services)

如现有的日志集/主题不合适，您可以去控制台 [新建日志集](#)

选择完成后，启动服务，进入服务日志详情页，您也可以通过页面上方的超链接进入日志服务检索页面。

← pmml1234

基本信息 实例列表 监控信息 事件信息 **运行日志**

ac4s2j7w56kf2h7q-74877b8458-gdwmw ▾

实时

近24小时

重置

[前往日志分析服务](#)

```
1 + checkVariable SECRET_ID, AKIDQ5gpark7eozE7dgAZhVS-cwZh9DdMntLmK-7rcF0MQ6PyqQWNTekB-50wNg-hPWq
2 + '[' -z AKIDQ5gpark7eozE7dgAZhVS-cwZh9DdMntLmK-7rcF0MQ6PyqQWNTekB-50wNg-hPWq ]]
3 + checkVariable SECRET_KEY, 'i04SwxubRE4rzjK8pIjHKCVxPCqbqhwoY5KXrk1r9Q='
4 + '[' -z 'i04SwxubRE4rzjK8pIjHKCVxPCqbqhwoY5KXrk1r9Q=' ]]
5 + checkVariable BUCKET, ti-ems-1255502019
6 + '[' -z ti-ems-1255502019 ]]
7 + checkVariable ENDPOINT, cos.ap-beijing.myqcloud.com
8 + '[' -z cos.ap-beijing.myqcloud.com ]]
9 + checkVariable REMOTE_PATH, models/pmml/
10 + '[' -z models/pmml/ ]]
11 + checkVariable LOCAL_PATH, /data/model/m
```

3. 开启日志检索

若您选择投递至 TI-EMS 默认日志集和日志主题，系统将自动开启索引，若您选择投递至自定义日志集和日志主题，请保证您的日志主题已开启索引，相关操作请查看 [日志检索](#)。

更多有关日志服务的使用方式请查看 [日志服务产品文档](#)。

资源组管理

最近更新时间：2021-11-16 09:55:21

在 TI-EMS 中，集群资源会被分为不同的资源组进行隔离，在启动模型服务的时候，用户可以选择将集群资源部署在默认的公共资源组或者部署在用户独享的专用资源组中。部署在不同资源组的模型服务计费方式详见 [计费概述](#)。

资源组类型	说明
公共资源组	公共资源组提供公共的大规模计算集群，根据用户服务配置按需分配使用。按资源规格、使用时长及实例数计费。
专用资源组	提供独享的计算资源，可用于模型部署。专用资源组不与其他用户共享，更加高效。购买资源节点后即开始计费，目前支持包年包月和按量计费两种计费模式。使用专属资源组部署上线服务，不再另行收费。

公共资源组

若用户选择在公共资源组部署服务，可自由选择占用的资源：CPU 配置的部署单位为 2Gquota 和 4Gquota，GPU 配置部署的单位为 P4quota。

部署单位说明：

- 2Gquota 为1核 CPU，2G 内存。
- 4Gquota 为1核 CPU，4G 内存。
- P4quota 为1核 CPU，5G 内存，0.25P4卡。

选定想要部署的资源类型以后，可指定 Quota 数量，1个实例副本的 2Gquota 和 4Gquota 的数量设置范围为1 - 10，1个实例副本的 P4quota 的数量设置范围为1 - 4（即单副本最大 GPU 配置为1P4卡）。

专用资源组

TI-EMS 的用户可以创建专属于自己的资源组，托管给 TI-EMS，用户可将模型服务部署在专属资源组上，不与其他用户共用资源。

专用资源组优势

- 对于业务稳定，需要长期使用的服务来说，使用预付费专用资源组可以为您节省总消耗成本。
- 用户可以在专用资源组中使用自定义运行环境，可灵活部署更多种类型的模型。
- 专用资源组支持和用户自己的 VPC 打通，服务调用更加高效。

专用资源组使用方式

新建资源组

进入资源组管理页面，单击**新建资源组**，输入资源组名称，选择资源组所在地域，单击**确定**，完成资源组创建。

新建资源组



资源组名称

名称为中文、字母、数字、下划线的组合，首字符必须是中文或字母，1-20位且不能为空

地域



创建资源组时不需要购买节点资源，资源组具体操作请见[产品文档](#)

资源组增加节点

资源组内手动增加节点入口有2处：

1. 在资源组列表，选择需要扩展节点的资源组，单击**增加节点**，进入节点购买页。

ID/名称	节点数	在线推理	批处理作业	地域	CPU(Core)	MEM(GB)	GPU(卡)	操作
rsg-k76dqtr rsg-k76dqtr	2	0	0	北京	24	48	0	增加节点 一键续费 删除
rsg-4cj6hkv6 rsg-4cj6hkv6	1	0	0	北京	8	16	0	增加节点 一键续费 删除
rsg-m6pb7hhw jimmy-test	0	0	0	北京	8	16	0	增加节点 一键续费 删除

2. 单击**资源组名称**，进入资源组节点详情页，单击**增加节点**，进入节点购买页。

增加节点		一键续费	CPU 24 核	MEM 48 GB			
ID	节点名称	节点类型	可用区	计费模式	状态	已分配/总资源	所属伸缩组
1	tiems-shmb...	16核32G	北京三区	按量计费	正常	CPU: 0.302/16 核 MEM: 0.46/32 GB	asg-7k2j66zc
2	tiems-q5dqjflp	8核16G	北京四区	按量计费	正常	CPU: 0.302/8 核 MEM: 0.46/16 GB	-

资源组购买方式请详见 [购买方式](#)。

⚠ 注意：

可能会因为当前地域或可用区资源不足等原因，导致节点部署失败，退款订单会自动创建，已支付的钱款会原路返回。

节点购买成功后，在 TI-EMS 相关资源组节点详情页就可以看到当前资源组下所有节点信息，在部署模型时也可以选择该专用资源组进行部署。

资源组节点详情

在资源组列表页，用户可以修改资源组名称，单击资源组名称，可进入查看资源组下各个节点的状态和资源使用情况。

资源组节点状态分为部署中、部署失败、正常、已过期四种类型。部署成功的节点状态显示为正常，可在此资源组正常启动模型服务（剩余资源充足的条件下），状态为部署中、部署失败或者已过期的节点不可部署模型服务，为了保证您的服务运行正常，请时刻关注您的包年包月节点到期状况，关于节点的过期说明请查看 [欠费说明](#)。

用户可对资源组进行一键续费和增加节点的操作，对包年包月资源组节点可进行节点续费、设置/取消自动续费的操作，对按量计费资源组节点可进行销毁操作。

资源组自动伸缩

对于访问量快速变化的业务，您可以选择创建伸缩组来满足资源的自动扩展需求。

创建伸缩组

1. 在资源组列表页点击资源组名称，进入资源组详情页-伸缩组页，单击**新建伸缩组**，进入伸缩组创建页面。

伸缩组配置

起始节点数

最小伸缩数

最大伸缩数

在设定的节点范围内自动调节，不会超出该设定范围

2. 输入以下信息：

- 伸缩组名称：输入1-20字符数量的伸缩组名称。
- 伸缩组的启动配置：选择伸缩组想要绑定的节点类型，目前伸缩组支持所有 TI-EMS 专用资源组按量计费的节点，节点规格和定价信息请查看 [定价说明](#)。
- 起始节点数：定义了伸缩组启用初始的节点数量。
- 最大伸缩数：定义了伸缩组内节点的最小数量；若当前节点数量小于最小伸缩数，TI-EMS 伸缩组将自动添加节点，使其等于最小伸缩数。
- 最小伸缩数：定义了伸缩组内节点的最大数量；若当前节点数量大于最大伸缩睡，TI-EMS 伸缩组将自动减少节点，使其等于最大伸缩数。

3. 单击**确定**，完成创建。

伸缩组操作

1. 启用/停用伸缩组

当伸缩组创建完成以后，需要启用伸缩组。单击**伸缩组列表>操作>启用**，伸缩组将按照平台伸缩规则进行节点扩展和销毁。

2. 更新伸缩组

当资源组配置的节点数量无法满足业务需求时，您可以单击**伸缩组列表>操作>更多>更新**，进行资源组伸缩数范围的更新，更新完成以后伸缩组将按照新的节点数量范围进行扩展与销毁。

3. 删除伸缩组

单击**伸缩组列表>操作>更多>删除**，即可删除伸缩组，删除伸缩组的前提是伸缩组内所有节点均已销毁。

4. 查看资源组节点列表

单击伸缩组名称，进入伸缩组详情 > 节点列表，当访问量增加，伸缩组会自动扩展节点，伸缩组内扩展的节点可在此页面进行查看。伸缩组内扩展的节点可以手动销毁。

5. 查看资源组伸缩记录

单击伸缩组名称，进入伸缩组详情 > 伸缩记录，您可查看伸缩组内节点的历史扩展与销毁记录。

TI-EMS 资源组伸缩规则

1. 扩容规则

当资源组中出现因为缺少可用资源而无法调度的容器实例时，将触发自动扩容策略，并尝试扩容节点来运行这些实例。用于判断资源是否充足的标准是资源组中剩余可用资源不满足服务实例的配置值。

当资源组下有多个伸缩组时，系统采用 least-waste 算法进行扩容操作，即选择调度后剩余资源更少的伸缩组进行扩容。

2. 缩容规则

TI-EMS 规定的缩容规则是当资源组中节点空闲资源较多时，将触发缩容。当节点上所有 Pod 的 CPU 或者内存占用资源/可分配资源的比值小于50%时，并且持续10分钟，该节点就会尝试缩容。同时，只有该节点上所有 Pod 都能够成功调度到其他节点上时，Pod 才会被驱逐从而缩容成功。

包年包月资源组节点续费

1. 资源组一键续费

用户可在 TI-EMS 控制台或腾讯云费用中心的续费管理页面对资源组进行一键续费操作，一键续费可将资源组下所有包年包月节点进行批量续费。关于一键续费的详细说明详见 [续费说明](#)。

2. 节点续费

用户可在 TI-EMS 控制台或腾讯云费用中心的续费管理页面对包年包月节点进行续费操作，关于节点续费的详细说明详见 [续费说明](#)。

3. 节点设置/关闭自动续费

用户可在 TI-EMS 控制台或腾讯云费用中心的续费管理页面对包年包月节点进行自动续费管理，关于节点自动续费的详细说明详见 [续费说明](#)。

资源组类型

TI-EMS 所支持的资源组节点类型和定价信息，请参见 [定价说明](#)。