

云托管 CloudBase Run 部署指南



腾讯云

【 版权声明 】

©2013–2025 腾讯云版权所有

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

【 商标声明 】



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

【 服务声明 】

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

【 联系我们 】

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

文档目录

部署指南

部署概述

创建环境

创建服务

更新服务

回滚服务

删除服务

删除环境

CLI 工具

云开发中云托管部署指南

新建服务

部署服务

更新或回滚服务

删除服务

部署指南

部署概述

最近更新时间：2024-08-05 15:23:11

本文介绍开始部署前您需要了解的一些基本事项。

使用限制

云托管暂不支持以下场景：

- 当前仅支持 HTTP 协议，通过域名调用服务。不支持 tcp/udp/mqtt 等其他协议。
- 不支持通过公网 IP 访问服务。
- 不支持部署 Docker Compose。
- 不支持接入 WAF。
- 不支持内网域名。
- 不支持一个服务开启多个监听端口。
- 不支持用来部署数据库/Redis 等有状态服务。

概念解释

环境

- 环境是所有资源的依托，使用云托管必须先创建环境。
- 环境之间所有资源都互相隔离，环境内资源之间可以通过内网调用，不同环境的资源之间只能使用公网调用。
- 不同环境独立计费并出账单。
- 不同环境无法共享资源包和套餐。
- 账号欠费时，所有环境都将被冻结，欠费超过7天，所有环境都将被删除，无法恢复。

用法推荐

- 用法一：根据用途区分环境，例如区分为测试环境、预发环境、生产环境等。
 - 服务的管理、成员权限分配独立。例如生产环境权限控制可严于测试环境。
 - 可根据实际用途配置不同级别的告警策略。例如生产环境告警条件更细致严格。
 - 独立控制各环境的资源消耗、费用与预算。例如测试环境可以采用更小规格或更少数量的实例。
- 用法二：根据不同上层用户区分环境。例如 SaaS 类业务可以为不同用户分配对应的独立环境。
 - 使不同用户的服务间逻辑隔离互不影响。
 - 每个用户实际使用资源基于不同环境分别计费。
 - 每个用户环境有独立的成员权限。

服务

服务是提供业务功能的主体，从现实感知角度上，一个服务对应一个域名，也就是调用方可以通过固定的域名访问到这个服务，不受服务更新迭代的影响。

服务的每次更新，都会产生一个新的版本，因此服务下可能包含多个版本。但服务在线上运行时只能有一个版本（一般是最新的版本）接入流量，响应请求。

特殊情况：在服务升级的过程中，如果需要灰度测试，可以同时有新旧两个版本都接入流量，这就是灰度发布。服务升级结束后，必须将所有流量都接入到新版本，或者全部回退到旧版本。

您需要根据业务需求，自行划分服务范畴，可以是一个功能（微服务的一个模块），也可以是几个功能的复杂组合（单一服务）。业务划分为不同的服务之后，不同服务之间就可以各自独立迭代，减少互相耦合影响；也可以各自独立扩缩容，更加精细地控制成本。

服务创建时，系统会分配一个公网域名，后续可绑定您购买并备案好的自定义域名。服务不支持通过 IP 访问。

版本

版本是由**镜像启动的容器实例**和相关配置组成的集合体，切实拥有运行的资源，可以直接发布到线上接入流量。每次更新服务，都会创建出新的版本代替旧版本。

一个服务下有多个版本，版本号/部署 ID 会自动递增，不能自定义或修改。

支持最多2个版本有流量（服务灰度过程中），其余历史版本不能配置流量，只可用于回滚。

实例

每个版本根据当前流量负载情况+在服务设置中指定的实例最小最大值，会对应0~N个实例（容器）。可以理解为他们都是彼此的“复制”和“分身”，运行着完全一模一样的业务逻辑，可以分担处理请求。服务负载高时需要的实例数量多，服务负载低时需要的实例数量少，服务没有任何访问时，甚至可以不保有任何实例（等再次有访问时重新创建实例）。

创建环境

最近更新时间：2024-11-07 16:27:21

⚠ 注意：

本篇文档适用于云开发新版控制台。

操作步骤

步骤1：角色授权（仅首次创建环境）

首次创建环境，需要给云托管授权以下角色，获取操作 EKS\TKE\TCB\CLB\CBS\CLS 等相关云资源的权限。请按页面提示跳转去同意授权（可能需要多次授权）。

步骤2：选择地域

登录 [云托管独立控制台](#)，在服务概览页顶部，切换所需地域。目前支持北京、上海和广州。

⚠ 注意：

已经创建好的环境，不支持更换地域。

步骤3：创建环境

单击[创建环境](#)，填写表单。表单字段信息详解，请参见 [环境设置](#)。

创建服务

最近更新时间：2025-03-17 12:01:32

⚠ 注意：

本篇文章适用于云开发新版控制台。

操作步骤

步骤1：选择环境

登录 [云托管独立控制台](#)，在服务概览页顶部选择对应的环境。进入 [云开发平台-云托管](#)，填写表单并准备好需要部署的服务代码。表单字段信息详解，请参见 [服务配置](#)。

⚠ 注意：

已经创建好的服务，不支持更换所在环境。

步骤2：开始部署

查看部署进度及日志，等待部署完成。首次部署成功后，会自动全量发布。

可选步骤：取消部署

取消部署无法立刻终止底层的部署任务，但会拦截本次版本的最终上线，同时不阻塞立刻发起下一次更新服务。当前创建中的服务不会因为取消部署而被删除。

步骤3：验证服务

云托管为每个服务分配了一个默认公网域名，可用于测试和验证服务。**默认域名有性能限制，请勿用于线上生产，不保证 SLA 及相关补偿。**

在服务概览页中的服务卡片或服务详情页的基本信息中，单击默认域名，即可访问服务并进行测试。也可以使用 POSTMAN 等工具进行接口测试。

步骤4：设置日志服务

因为容器无法持久化存储文件，且实例因为弹性伸缩处在不断变化之中，登录具体容器实例去查看日志极为繁琐不便。云托管采取的是将日志通过 Loglistener 组件采集出来，投递到腾讯云日志服务 CLS 中持久保存并查看。在 [服务详情页](#) > [服务设置](#) > [运行配置](#) > [日志采集类型](#) 中，修改为自定义，然后继续填写日志服务所需的日志集、日志主题、日志格式，保存后将触发重新部署，成功后生效。日志采集类型为无则将不采集日志。

步骤5：绑定自定义域名

1. 服务测试完毕需要上线生产前，前往 [服务详情页](#) > [服务设置](#) > [访问管理](#) 中，单击 [添加域名](#)。

- 域名：目标自定义网域名称（Domain Name），由一串用点分隔的字符组成。域名在云托管内用于访问服务，一个域名只能绑定一个服务，**域名一定需要完成域名备案，否则无法访问**；云托管绑定域名也支持在其他云服务厂商中注册的域名。
 - HTTPS：访问方式目前支持 HTTP、HTTPS 2种方式，默认使用 HTTPS。
 - 证书：HTTPS 访问需要上传证书，可前往购买证书的平台下载证书签名及私钥，证书文件分证书和私钥，一般建议传 `nginx` 等类型的文本证书，不支持二进制可执行证书。
2. 单击**确定**，即可将自定义域名绑定到目标服务上，并在列表中生成一条绑定信息，提供**编辑及解绑**能力。绑定自定义域名后，域名列表中会返回 CNAME 信息；可前往 DNS 服务提供商（例如 [腾讯云域名](#)）添加 CNAME 记录，将域名指向云托管返回的服务 CNAME 域名；如在 DNSPod 中配置，请参见 [CNAME 配置指引](#)。

ⓘ **说明：**

证书过期更新，直接**编辑**重新上传证书，不需要删除重新配置。

步骤6：配置监控 Dashboard 及告警（可选）

在**服务详情页 > 监控**中，云托管提供了一定的基础监控视图。如需更高阶的监控功能，请前往 [腾讯云可观测平台控制台 Dashboard 列表](#)。

如需根据某个监控指标配置告警，请前往 [腾讯云可观测平台控制台告警配置](#)。

更新服务

最近更新时间：2024-11-07 16:27:21

⚠ 注意：

本篇文章适用于云开发新版控制台。

代码更新操作步骤

步骤1：进入服务详情页面

登录 [云托管独立控制台](#)，单击需要更新的服务，进入服务详情页。

步骤2：开始更新

⚠ 注意：

同一个服务无法有多个更新操作并行（代码更新、配置更新）。

单击**更新服务**，填写表单。支持全量更新或灰度发布两种更新策略。更多字段详细信息，请参见 [服务设置](#)。

- **全量更新**：部署成功后，立刻将全部流量切换到新版本。半小时后，旧版本因无流量访问而下线。此过程中用户感知是平滑的，服务不会中断。
- **灰度发布**：部署成功后，回到**服务详情页 > 概览**，在服务部署卡片或部署记录中，单击**配置流量**，手动调节新旧版本之间的流量百分比，将用户请求按比例随机分配至新旧版本。此配置流量过程可以反复多次，过程中用户感知是平滑的，服务不会中断。

📌 说明：

通过监控和日志观测新版本的运行情况，确认新版本一切就绪，在**配置流量**中选择100%，结束灰度过程，半小时后，旧版本因无流量访问而下线。如果新版本有问题，单击**放弃灰度**，新版本下线。

步骤3：验证服务

云托管为每个服务分配了一个默认公网域名，可用于测试和验证服务。**默认域名有性能限制，请勿用于线上生产，不保证 SLA 及相关补偿。**

在服务概览页中的服务卡片或服务详情页的基本信息中，单击默认域名，即可访问服务并进行测试。也可以使用 POSTMAN 等工具进行接口测试。

可选步骤：取消部署

取消部署无法立刻终止底层的部署任务，但会拦截本次版本的最终上线，同时不阻塞立刻发起下一次更新服务。

配置更新操作步骤

步骤1: 进入服务详情页面

登录 [云托管独立控制台](#)，单击需要更新的服务，进入服务详情页。

步骤2: 编辑服务设置

⚠ 注意:

服务正在更新中时，无法编辑配置。

进入服务设置页面，编辑代码配置、运行配置或自动伸缩配置。保存后会用新的配置触发一次部署，部署成功后，立刻将全部流量切换到新版本。半小时后，旧版本因无流量访问而下线。此过程中用户感知是平滑的，服务不会中断。

步骤3: 验证服务

云托管为每个服务分配了一个默认公网域名，可用于测试和验证服务。**默认域名有性能限制，请勿用于线上生产，不保证 SLA 及相关补偿。**

在服务概览页中的服务卡片或服务详情页的基本信息中，单击默认域名，即可访问服务并进行测试。也可以使用 POSTMAN 等工具进行接口测试。

可选步骤: 取消部署

取消部署无法立刻终止底层的部署任务，但会拦截本次版本的最终上线，同时不阻塞立刻发起下一次更新服务。

回滚服务

最近更新时间：2024-11-07 16:27:21

⚠ 注意：

本篇文档适用于云开发新版控制台。

适用场景

服务正常运行中，发现紧急故障或问题，需要快速还原到更早之前的版本。

操作效果

- 回滚不需要重新上传代码并部署。
- 确定执行回滚后，会先冷启动所选旧版本（有一定冷启动耗时，取决于旧版本的启动性能），等旧版本就绪后，再将所有用户请求转移到指定的旧版本上。
- 被放弃的版本（回滚之前的线上版本）失去流量后，下线缩容。
- 服务的所有设置（代码配置、运行配置、自动伸缩）固化在了旧版本对应镜像中，因此实际线上运行的都将被还原到旧版本发布时的值，可能与此时在**服务设置** TAB 中看到的值有所不同。再次更新服务，仍将使用**服务设置** TAB 中看到的值。

操作步骤

步骤1：进入服务详情页面

登录 [云托管独立控制台](#)，单击需要回滚的服务，进入服务详情页。

步骤2：执行回滚

单击回滚，选择要回滚到的版本，然后执行。

ⓘ 说明：

不支持向前回滚（版本号高于当前线上版本），如有需要请走更新服务流程。

步骤3：验证服务

在服务详情页中，可以看到当前100%流量的已经是回滚目标 ID。

云托管为每个服务分配了一个默认公网域名，可用于测试和验证服务回滚效果是否符合预期。**默认域名有性能限制，请勿用于线上生产，不保证 SLA 及相关补偿。**

在服务概览页中的服务卡片或服务详情页的基本信息中，单击默认域名，即可访问服务并进行测试。也可以使用 POSTMAN 等工具进行接口测试。

删除服务

最近更新时间：2024-11-07 16:27:21

⚠ 注意：

本篇文章适用于云开发新版控制台。

适用场景

服务永久下线，不再运行。

操作效果

- 服务的默认域名永久失效，无法再被复用，调用后都会报错。
- 服务绑定的自定义域名调用会报错，除非重新绑定到其他服务上。
- 服务不再产生用量，也不会再产生任何费用。（服务资源消耗和计费是秒级的，删除之后仍会最后结算一次费用）。
- 服务一经删除，不可恢复，再次创建的服务哪怕同名，也不被认为是同一个服务，默认域名也会变化。
- 删除服务，对应的服务镜像也将删除不再继续保存在镜像仓库中，但不会对服务对应的代码仓库或代码包产生任何影响。您可以用相同的代码仓库或代码包再次创建新的服务。

操作步骤

步骤1：进入服务详情页面

登录 [云托管独立控制台](#)，单击需要删除的服务，进入服务详情页。

步骤2：进入服务设置

进入服务设置页面，在最底部单击**删除服务**。

删除环境

最近更新时间：2024-11-07 16:27:21

⚠ 注意：

本篇文章适用于云开发新版控制台。

适用场景

- 创建环境时 VPC 或地域选择错误，需要重新创建，配置错误的环境弃用。
- 注销关闭云托管，不再使用。

操作效果

- 环境下所有服务下线失效，无法调用，不可恢复。
- 环境一经删除不可恢复和找回。
- 环境订阅的套餐将按剩余有效期退款。
- 环境内若有未使用的资源包，会默认退款。
- 因服务的按量计费是精确到秒，因此删除环境后，其下所有服务还会进行最后一次费用结算。
- 注销关闭云托管，需要删除账号下所有的环境。

操作步骤

步骤1：进入环境设置

登录 [云托管独立控制台](#)，左侧菜单单击进入环境设置页面。

步骤2：删除环境

单击**删除环境**。删除环境之前，请确认已删除环境下所有服务。

CLI 工具

最近更新时间：2024-11-07 16:27:21

云托管推出 CLI 工具，帮助开发者能够在本地或者自定义 CI/CD 中快速进行版本创建和其他操作。

⚠ 注意：

CLI 工具仅限云托管新版控制台的环境使用，云开发中的云托管无法使用。

安装

CLI 工具安装前需要安装 `npm`，具体请参见 [Downloading and installing Node.js and npm](#)。

```
npm install -g @cloudbase/cli
```

如果 `npm install -g @cloudbase/cli` 失败，您可能需要 [修改 npm 权限](#)，或者以系统管理员身份运行：

```
sudo npm install -g @cloudbase/cli
```

如果安装过程没有错误提示，一般就是安装成功了。下面，我们可以继续输入命令：

```
tcb -v
```

如果看到输出版本号，说明已经安装成功。

使用代理

使用 CLI 工具时，需要您的终端能够访问公网。如果您的终端无法直接访问公网，您可以设置 HTTP 代理使 CLI 能够正常使用。CLI 会读取 `http_proxy` 或 `HTTP_PROXY` 环境变量，自动设置网络代理服务。

例如，您可以在终端中运行以下命令，设置 CLI 通过 `http://127.0.0.1:8000` 的代理服务访问网络：

```
export HTTP_PROXY=http://127.0.0.1:8000
```

⚠ 注意：

`http://127.0.0.1:8000` 只是一个示例，请勿直接使用，具体的代理配置根据您的网络配置决定。

上面的命令只是临时设置，当您关闭终端后，代理会自动失效，下次开启终端后需要重新设置。如果您需要一直通过代理访问公网，可以把命令加入到终端的配置文件中。

查看服务列表

查看指定环境下的所有服务列表：

```
tcb run service:list [OPTIONS]
```

参数信息：

OPTIONS	
-h, --help	查看帮助信息
-e, --envId	环境ID, 必填
-s, --serviceName	服务名
--json	以 JSON 形式展示结果
-r	地域参数, 不填默认操作上海地域环境。可取值：上海 (sh), 广州 (gz), 北京 (bj)

在指定环境下创建服务

创建服务并发布一个版本：

```
tcb run service:create [OPTIONS]
```

参数信息：

OPTIONS	
--noConfirm	发布前是否跳过二次确认
--override	缺省的参数是否沿用旧版本配置
-e, --envId <envId>	环境 Id, 必填
-s, --serviceName <serviceName>	服务名, 必填
-r	地域参数, 不填默认操作上海地域环境。可取值： 上海 (sh), 广州 (gz), 北京 (bj)
--path <path>	本地代码根目录
--cpu <cpu>	单一实例cpu规格, 默认0.5
--mem <mem>	单一实例内存规格, 默认1
--minNum <minNum>	最小副本数, 默认0
--maxNum <maxNum>	最大副本数, 默认50, 不能大于50
--policyDetails <policyDetails>	扩缩容配置, 格式为条件类型=条件比例(%), 多个条件之间用&隔开, 内存条件为mem, cpu条件为cpu, 默认内存>60% 或 CPU>60%, 即cpu=60&mem=60
--customLogs <customLogs>	日志采集路径, 默认stdout
--envParams <envParams>	环境变量, 格式为xx=a&yy=b, 默认为空

```

--containerPort <containerPort> 监听端口，必填
--remark <remark> 版本备注，默认为空
--targetDir <targetDir> 目标目录
--dockerfile <dockerfile> Dockerfile文件名，默认为 Dockerfile
--custom_image <custom_image> 容器镜像仓库企业版（TCR）镜像 URL
--json 以 JSON 形式展示结果
-h, --help 查看命令帮助信息

```

更新发布服务

指定环境和已存在的服务，发布新的更新：

```
tcb run service:deploy [OPTIONS]
```

参数信息：

```

OPTIONS
--noConfirm 发布前是否跳过二次确认
--override 缺省的参数是否沿用旧版本配置
-e, --envId <envId> 环境 Id，必填
-s, --serviceName <serviceName> 服务名，必填
-r 地域参数，不填默认操作上海地域环境。可取值：
上海（sh），广州（gz），北京（bj）
--path <path> 本地代码根目录
--cpu <cpu> 单一实例cpu规格，默认0.5
--mem <mem> 单一实例内存规格，默认1
--minNum <minNum> 最小副本数，默认0
--maxNum <maxNum> 最大副本数，默认50，不能大于50
--policyDetails <policyDetails> 扩缩容配置，格式为条件类型=条件比例（%），多个条件之间用&隔开，内存条件为mem，cpu条件为cpu，默认内存>60% 或 CPU>60%，即cpu=60&mem=60
--customLogs <customLogs> 日志采集路径，默认stdout
--envParams <envParams> 环境变量，格式为xx=a&yy=b，默认为空
--containerPort <containerPort> 监听端口，必填
--remark <remark> 版本备注，默认为空
--dockerfile <dockerfile> Dockerfile文件名，默认为 Dockerfile
--custom_image <custom_image> 容器镜像仓库企业版（TCR）镜像 URL
--json 以 JSON 形式展示结果
-h, --help 查看命令帮助信息

```

更新或新建服务

在指定的环境中，新增或更新服务（如果服务不存在会创建）。

```
tcb run deploy [OPTIONS]
```

参数信息：

OPTIONS

<code>--noConfirm</code>	发布前是否跳过二次确认
<code>--override</code>	缺省的参数是否沿用旧版本配置
<code>-e, --envId <envId></code>	环境 Id, 必填
<code>-s, --serviceName <serviceName></code>	服务名, 必填
<code>-r</code>	地域参数, 不填默认操作上海地域环境。可取值： 上海 (sh), 广州 (gz), 北京 (bj)
<code>--path <path></code>	本地代码根目录
<code>--cpu <cpu></code>	单一实例cpu规格, 默认0.5
<code>--mem <mem></code>	单一实例内存规格, 默认1
<code>--minNum <minNum></code>	最小副本数, 默认0
<code>--maxNum <maxNum></code>	最大副本数, 默认50, 不能大于50
<code>--policyDetails <policyDetails></code>	扩缩容配置, 格式为条件类型=条件比例(%), 多个条件之间用&隔开, 内存条件为mem, cpu条件为cpu, 默认内存>60% 或 CPU>60%, 即 cpu=60&mem=60
<code>--customLogs <customLogs></code>	日志采集路径, 默认stdout
<code>--envParams <envParams></code>	环境变量, 格式为xx=a&yy=b, 默认为空
<code>--containerPort <containerPort></code>	监听端口, 必填
<code>--remark <remark></code>	版本备注, 默认为空
<code>--targetDir <targetDir></code>	目标目录
<code>--dockerfile <dockerfile></code>	Dockerfile文件名, 默认为 Dockerfile
<code>--custom_image <custom_image></code>	容器镜像仓库企业版 (TCR) 镜像 URL
<code>--json</code>	以 JSON 形式展示结果
<code>-h, --help</code>	查看命令帮助信息

更新服务配置

指定环境和服务，更新服务的基础配置：

```
tcb run service:config [OPTIONS]
```

参数信息：

OPTIONS

<code>-e, --envId <envId></code>	环境 Id, 必填
--	-----------

```
-s, --serviceName <serviceName> 服务名, 必填
-r 地域参数, 不填默认操作上海地域环境。可取值:
上海 (sh), 广州 (gz), 北京 (bj)
--cpu <cpu> 单一实例cpu规格, 默认0.5
--mem <mem> 单一实例内存规格, 默认1
--minNum <minNum> 最小副本数, 默认0
--maxNum <maxNum> 最大副本数, 默认50, 不能大于50
--policyDetails <policyDetails> 扩缩容配置, 格式为条件类型=条件比例(%), 多
个条件之间用&隔开, 内存条件为mem, cpu条件为cpu, 默认内存>60% 或 CPU>60%, 即
cpu=60&mem=60
--customLogs <customLogs> 日志采集路径, 默认stdout
--envParams <envParams> 环境变量, 格式为xx=a&yy=b, 默认为空
--json 以 JSON 形式展示结果
-h, --help 查看命令帮助信息
```

云开发中云托管部署指南

新建服务

最近更新时间：2024-11-07 16:27:22

⚠ 注意：

本篇文档仅适用于云开发中的云托管。

操作流程

步骤1：新建服务

在 [云开发中的云托管控制台](#) 的服务列表页面，选择对应的环境，单击**新建服务**或**新建此私有网络下的服务**。

步骤2：配置服务信息

填写新建服务所需的所有信息。字段详解请参见 [服务配置说明](#)。

字段说明：

● 新建服务：

- 当前环境首个服务：**云托管网络**字段默认选中**系统推荐**。
- 当前环境非首个服务：**云托管网络**字段默认选中**选择已有私有网络**。
 - **私有网络**字段默认选中上一个服务所在 VPC。
 - **子网**字段默认选中上一个服务所在子网。

⚠ 注意：

每个账号都存在 VPC 总数配额，在无特殊的网络需求前提下，建议您尽量将多个服务部署在同一个 VPC 内以节约配额。

● 新建此私有网络下的服务：

- **云托管网络**字段默认选中**选择已有私有网络**。
 - **私有网络**字段默认选中您所选的 VPC。
 - **子网**字段需要手动选择。

步骤3：提交

单击**提交**，如果部署成功，便可以看到如下弹框：



单击**新建版本**可立刻开始新建版本并部署，详情请参见 [部署服务](#)。

使用限制

同一账号下，环境 + VPC 的不同组合最多存在20个。以下几种情况可能导致超出配额限制，引发报错无法创建新的服务。**每次创建服务时云托管网络选择系统创建则会新建一个 VPC：**

- 同一环境下，使用了20个不同 VPC 部署了多个服务。如果新建服务时选择了第21个未使用过的 VPC/**云托管网络选择了系统创建**。
- 在20个不同环境中，使用同一个 VPC 部署了多个服务。如果在第21个环境中新建服务，或在存量20个环境中新建服务时选择了当前环境未使用过的 VPC/**云托管网络选择了系统创建**。
- 以上两种情况的混合，使得环境 + VPC 的不同组合数超过20个。

部署服务

最近更新时间：2024-11-07 16:27:22

⚠ 注意：

本篇文章仅适用于云开发中的云托管。

操作步骤

步骤1：进入服务详情页面

登录 [云开发中的云托管控制台](#)，选择您需要部署的服务，单击服务名称进入服务详情页面。

📌 说明：

如果您还没有任何服务，请先单击[新建服务](#)，详情参见 [新建服务](#)。

步骤2：新建版本

单击[新建版本](#)，在新建版本窗口中，继续填写版本所需配置信息，详情请参见 [版本配置说明](#)。

步骤3：开始部署

填写完版本配置信息后，单击[开始部署](#)，部署成功则状态变为正常。若有报错，会变为具体的错误状态。

- 流量策略选择了[部署完成后自动开启100%流量](#)，则流量会从0%变为100%。
- 流量策略选择了[部署完成后保持流量为0稍后再手动调整流量](#)，或者选择了[部署完成后自动开启100%流量但还想再次调整流量](#)，可以进行手动配置，详情请参见 [流量配置](#)。

步骤4：访问服务

完成了部署和流量配置后，您可能希望能够快速访问自己的服务查看效果。云托管自动为您的服务分配了一个默认域名，您可以直接单击[访问服务](#)，通过这个默认域名访问您的服务页面。

- 配置流量时选择的模式为[按百分比](#)，则单击[访问服务](#)将按您设定的百分比概率随机访问某个版本。
- 配置流量时选择的模式为[按 URL 参数](#)，则单击[访问服务](#)产生的是一个无参数的 HTTP 请求，会访问到您设置的默认版本上。您可自行构建含参数的请求进一步验证。

📌 说明：

单击[访问服务](#)产生的即为真实业务流量。因此，如果版本采用的是低成本模式，会触发实例个数从0扩容到1，伴随最长30秒的冷启动延迟以及这1个实例产生的资源消耗和费用。

使用限制

- 版本名由系统自动生成，格式为**服务名 + 序号**，序号按照创建顺序依次递增，不支持修改。例如 testservice-001, testservice-002 等。
- 不再使用的版本可以手动删除，但不会影响新建版本的序号。例如，删除掉版本 testservice-002，再次新建的版本仍然会是 testservice-003。

资源消耗说明

无论采用哪种副本模式，**部署过程本身都会产生一定的资源消耗。**

低成本模式

虽然副本个数最小值为 0（无流量时缩容到 0 个实例不产生资源消耗和费用），但在部署过程中会先创建出 1 个实例，待部署成功后再触发缩容到 0。

示例

副本模式为**低成本**，规格为**1 核 1G**，流量策略为**部署完成后自动开启100%流量**，部署耗时 5 分钟，部署成功后一直无任何业务流量。因为缩容到 0 的观测期为半小时，故会产生 CPU 使用量 35（核 x 分钟）以及内存使用量 35（GiB x 分钟）。

高可用模式

部署开始即会创建出最小个数的实例，产生资源消耗。

示例

副本模式为**高可用**，规格为**1 核 1G**，最小副本个数为 10，流量策略为**部署完成后保持流量为0稍后再手动调整流量**，部署耗时 5 分钟，则会产生 CPU 使用量 50（核 x 分钟）以及内存使用量 50（GiB x 分钟），即便部署成功后一直无任何业务流量，也将持续消耗资源。

更新或回滚服务

最近更新时间：2024-11-07 16:27:22

⚠ 注意：

本篇文章仅适用于云开发中的云托管。

云托管目前支持两种更新和回滚方式：

对比项	新建版本	编辑原版本配置并重新部署（不推荐）
更新前后内网 IP 不变	否，每个版本 IP 不同	是，版本不变则 IP 不变
调用方式	通过域名调用	通过域名调用或通过内网 IP 调用（限同一 VPC 内）
回滚	支持，流量导回旧版本即可回滚	支持，用历史配置重新部署实现回滚
灰度发布	支持，详情请参见 流量配置	不支持
多版本并行	支持	不支持

方式一：新建版本（推荐）

每次更新服务都新建一个版本，部署成功后逐步将流量从旧版本切换到新版本。通过域名调用不强依赖固定 IP，更符合云原生理念。

步骤1：进入服务详情页面

登录 [云开发中的云托管控制台](#)，选择您需要更新的服务，单击服务名称进入服务详情页面。

步骤2：新建版本

单击**新建版本**，填写更新所需的版本配置信息，详情请参见 [版本配置说明](#)。

根据需求选择合适的流量配置策略，例如：

- **灰度发布/蓝绿发布**：流量策略选择**部署完成后保持流量为 0**，稍后**手动配置流量**，待发布完成后，手动配置流量到新版本上。详情可参见 [流量配置](#)。
- **全量发布**：流量策略选择**部署完成后自动开启100%流量**，待发布完成后，即自动切换全部流量到新版本上。

步骤3：开始更新

单击**开始部署**。

步骤4：回滚服务

通过流量配置，将希望回滚的旧版本流量设置为100%，新版本流量设置为0。

方式二：原版本编辑配置并重新部署

不新建版本，在已有版本中编辑配置信息，然后重新部署。此方式下不支持灰度。

步骤1：进入服务详情页面

登录 [云开发中的云托管控制台](#)，选择您需要更新的服务，单击服务名称进入服务详情页面。

步骤2：编辑配置

选择需要更新的版本，在操作 > 更多下拉菜单中，选择编辑配置并重新部署。或单击目标版本进入版本详情页面，单击版本配置选项卡，在配置信息卡片右上角单击编辑配置并重新部署。

填写更新所需更新的版本配置信息，详情可参见 [版本配置说明](#)。

说明：

- 重新部署不支持重新选择流量策略，将沿用版本当前流量配置。如有需要，请在重新部署成功后手动配置流量。
- 若需更新镜像/代码，请单击更新镜像信息，并填写表单。提交新的镜像信息并不会立刻开始部署，需要等其余配置信息也编辑完毕后触发。

步骤3：重新部署

单击保存并重新部署，版本状态变为更新中。部署成功则状态变为正常，流量保持不变。若有报错，会变为具体的错误状态。

部署成功后，版本历史中会生成一个新的快照，记录此次部署前版本的详情，用于回滚。

步骤4：回滚

在版本历史中选择需要回滚的快照，单击查看详情对比当前版本和快照的差异，确认后单击回滚。

版本状态变为更新中。回滚成功则状态变为正常，流量保持不变。若有报错，会变为具体的错误状态。

风险说明

重新部署过程中，为保证业务持续运行，将采取滚动更新实例的方式进行升级。滚动更新可能带来以下风险：

- 重新部署需要所有实例更新成功才算成功，任意实例更新失败会自动回滚整个版本，导致版本更新不成功。
- 重新部署过程中，版本同时存在已更新的实例和尚未更新的实例，有一定概率不同请求访问同一版本时现象不一致。
- 重新部署过程中，因正在更新的实例无法提供服务，可能导致剩余实例无法承受当前业务流量而触发扩容条件，产生额外的实例及资源消耗。

删除服务

最近更新时间：2024-08-20 10:41:21

注意：

本篇文章仅适用于云开发中的云托管。

注意事项

删除服务为不可恢复操作，操作前，请先评估对您的业务的影响。

操作步骤

1. 登录 [云开发中的云托管控制台](#)，再按需切换到指定的环境。
2. 进入需要删除的服务详情。选择您需要删除的服务，单击服务名称进入服务详情页面。

新建服务	请输入关键字进行过滤			
服务名称	服务备注	创建时间	更新时间	操作
joantestcode	测试	2020-07-05 15:56:37	2020-07-05 15:56:37	删除
mytest	-	2020-07-28 18:30:40	2020-07-28 18:30:40	删除
testcase	-	2020-07-10 15:56:56	2020-07-10 15:56:56	删除
reptest	测试镜像仓库命名	2020-08-02 16:26:04	2020-08-02 16:26:04	删除
reptest2	-	2020-08-02 16:49:02	2020-08-02 16:49:02	删除

3. 删除服务下所有版本。对每个版本进行单击删除 > 完成操作，删除所有版本。



4. 在服务列表页面，选中服务，单击删除即可删除服务。

新建服务

Q
↻

服务名称	服务备注	创建时间	更新时间	操作
joantestcode	测试	2020-07-05 15:56:37	2020-07-05 15:56:37	删除
mytest	-	2020-07-28 18:30:40	2020-07-28 18:30:40	删除
testcase	-	2020-07-10 15:56:56	2020-07-10 15:56:56	删除
repotest	测试镜像仓库命名	2020-08-02 16:26:04	2020-08-02 16:26:04	删除
repotest2	-	2020-08-02 16:49:02	2020-08-02 16:49:02	删除