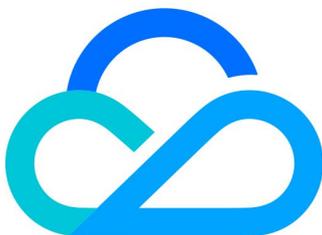


云托管 CloudBase Run 配置指南



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

配置指南

环境设置

服务设置

服务访问公网

子账号管理

监控告警设置

云开发中云托管配置指南

服务配置说明

版本配置说明

流量配置说明

配置指南

环境设置

最近更新时间：2024-06-05 17:19:31

⚠ 注意：

本篇文章仅适用于云托管新版控制台。

环境 ID

环境的唯一标识，创建成功后**不支持修改**。

环境 VPC

关于什么是私有网络（VPC）及子网，请参见 [私有网络](#)。**不支持更换环境所在 VPC。**

使用自有 VPC

开通环境时，选择高级定制版套餐，则需要使用您自有的 VPC。此 VPC 需要满足以下条件：

- 有至少2个子网，覆盖以下至少2个可用区（即所有子网不能都在同一个可用区）：
 - 广州三区，广州四区，广州六区，广州七区。
 - 上海二区，上海四区，上海五区。
 - 北京三区，北京四区。
- 子网的 CIDR 必须小于等于24，以保证充足的可用 IP 数。

授权管理

容器镜像仓库-企业版

需要先加入授权后，云托管才可获得对应的访问凭证，从而在部署服务时，选到对应的容器镜像仓库实例。

服务设置

最近更新时间：2024-06-05 17:19:31

⚠ 注意：

本篇文档仅适用于云托管新版控制台。

服务名称

环境下服务的唯一标识。创建成功后**不支持修改服务名称**。如果您有多个环境，不同环境之间服务可以重名。（例如：开发环境和测试环境可以同时存在服务"testservice"）。服务名只能包含数字、小写字母和“-”，且必须以小写字母开头。

部署方式

支持上传代码包和拉取镜像两种方式。

方式1：上传代码包

上传代码工程文件夹或 zip、jar、war 压缩包，系统解析成功后先将代码构建为镜像、将镜像推送到镜像仓库储存，然后基于该镜像进行部署。

- 此方式需要您自行编写 Dockerfile 并包含在代码文件夹或 zip 压缩包中，放置在目标目录的根目录下。
- 压缩包请注意不要直接压缩顶层文件夹，而是压缩文件夹内所有文件。否则会导致路径多一层而报错找不到 Dockerfile。

方式2：拉取镜像

无需系统提供从代码构建镜像的服务，由您自行采用任意方式构建好镜像后，手动将镜像上传至 [容器镜像服务（企业版）](#) 中，然后系统基于您选定的镜像进行部署。

- 使用[拉取镜像](#)方式部署前，请先到[环境设置](#)中授权需要使用的容器镜像服务实例，云托管才可获得访问凭证，从而拉取镜像。
- 容器镜像服务实例请尽量开放内网访问。如果走公网访问，拉取镜像将产生公网流量费用且速度较慢；公网/内网均不开启，则无法拉取镜像会导致部署失败。配置参见 [容器镜像服务-访问管理](#)。

端口

服务的监听端口，不支持多个。服务的调用方无需感知和使用此端口。监听端口必须为1~61000之间的整数且不能为9100。

环境变量

用户所需的环境变量，直接传入容器中。以 key value 的形式可配置多个。适合用来管理一些可变的配置，使得可以在不重新部署服务的情况下改变某些参数，例如数据库地址等。

目标目录

当代码包中包含不需要构建的文件时（例如同一个代码包中包含多个服务的代码，当前服务只用到其中一个子目录；或者需要排除 README 等文件时），可以将对应需要构建的子目录指定为目标目录，则不在目标目录中的文件会被忽略。

⚠ 注意：

Dockerfile 需要位于目标目录的根目录下，否则会报错。

Dockerfile 名称

默认为 Dockerfile。当同一份代码在不同场景时需要配套使用不同的 Dockerfile，就需要用不同名称区分 Dockerfile。

公网访问

关闭后，默认域名或自定义域名均不可访问。

默认域名

服务创建后即分配的系统默认域名，仅能支持测试使用，不承诺 SLA 及相关补偿，请勿用于线上生产。线上业务请绑定并使用自定义域名。

自定义域名

绑定您自行购买和已备案好的域名（不一定要在腾讯云购买及备案），一个服务可以绑定多个，但多个服务不能绑定同一个自定义域名。域名一定需要完成域名备案，否则无法访问。

访问方式目前支持 HTTP、HTTPS 2种方式，默认使用 HTTPS。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。

📌 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。

- 证书通常是以 `.crt` 或 `.pem` 为扩展名的文件，使用文本编辑器打开证书文件可以看到内容 `BEGIN CERTIFICATE` 开头，`END CERTIFICATE` 结尾。
- 私钥通常以 `.key` 或 `.pem` 等为扩展名的文件，使用文本编辑器打开证书文件可以看到内容 `BEGIN PRIVATE KEY` 开头，`END PRIVATE KEY` 结尾。

容器规格

每个容器实例的配置。自动扩容时，新创建的实例也将使用这个规格。同一个服务下所有运行中的容器实例规格都会保持一致。可选的容器规格范围由当前环境所选套餐决定。容器规格不会随着扩缩容而改变。

日志采集类型

选择**无**，则不采集日志（可以通过 webshell 登录容器实例查看日志文件，但因无法持久化存储，所以日志文件会在重启或扩缩容后消失）；选择**自定义**，则将日志采集到 [腾讯云日志服务 CLS](#) 中。

日志采集路径

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变。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。可设置采集目录或文件，支持设置多个路径，支持标准输出（stdout）以及 * 通配路径（例：/logs/*），使用 ,（半角逗号）分割，留空将采集标准输出。

⚠ 注意：

采集路径必须是最底层目录，其下不可以再包含文件夹，否则会采集失败。例如：日志如果存储在 /logs/today 中，日志采集路径配必须配置 /logs/today/*，而不能只配置到 /logs。

实例副本数

指当前服务在自动扩缩容时可达到的最大实例数及最小实例数。最小值下限为 0，最大值上限由当前环境所选套餐决定。

- 副本个数最小值为 0，当连续半小时**没有通过域名调用服务时**，版本将缩容到 0，不保留任何实例，不产生资源消耗。再次冷启动时，可能有一定的服务延迟。
- 虽然没有业务流量，部署过程中仍然会先产生一个实例，部署完成后再缩容到 0，因此**部署过程本身会产生一定的资源消耗**。
- 副本个数最小值不为 0，即便服务没有任何访问，仍会保持最小个数的实例数，也会**持续产生资源消耗**。

扩缩容条件

当达到某个条件时，云托管会自动会创建/删除实例，然后检测是否再次达到条件，如果满足条件则继续扩缩容，如此反复直至实例数量达到副本个数的最小值或最大值，或不再满足扩缩容条件时停止自动扩缩容。

目前支持 CPU 使用率或内存使用率作为扩缩容条件。

ⓘ 说明：

自动扩容到副本个数最大值后若仍不足以承载业务流量，即便再次达到扩缩容条件，也不会继续创建新的实例，可能导致您的业务受影响，请您评估好业务指标后合理设定副本个数最大值。

服务访问公网

最近更新时间：2024-11-07 16:27:22

⚠ 注意：

本篇文章仅适用于云托管新版控制台。

高级定制版套餐环境

使用高级定制版套餐的环境，因为使用的是您自有的 VPC，所以云托管中服务访问公网，需要在该环境的 VPC 中搭配 [NAT 网关](#)。每个 VPC 中只需要一个 NAT 网关即可支撑所有服务的公网访问诉求。

同时，这个 NAT 网关绑定的弹性 IP 配合您的路由表设置，决定您的云托管服务访问公网时的出口 IP。具体请参见 [NAT 网关相关文档](#)。

操作步骤

步骤1：确认环境所在 VPC

登录 [云开发新版控制台](#)，左侧菜单单击进入环境设置页面，获取环境 VPC 信息。

步骤2：购买 NAT 网关

前往 [NAT 网关控制台](#)，单击[新建购买 NAT 网关实例](#)。在网关配置中，请确保地域、私有网络，均与云托管的环境 VPC 一致。出带宽上限根据业务实际情况选择，弹性公网 IP 配置选择[新建弹性公网 IP](#)。



步骤3：配置路由表

根据 [配置指向 NAT 网关的路由](#)，将云托管环境 VPC 中的子网，逐个配置好路由策略。目的端如不限制（云托管服务可以访问任意公网地址），可以填写 0.0.0.0/0。

子账号管理

最近更新时间：2024-11-07 16:27:22

⚠ 注意：

本篇文档仅适用于云托管新版控制台。

适用场景

授权子账号操作云托管环境/服务，允许或限制子账号只能操作特定的云托管环境。

操作步骤

步骤1：用主账号完成云托管开通授权

用主账号登录 [云开发新版控制台](#)，按页面提示跳转去访问管理控制台同意授权（可能需要多次授权）。

📌 说明：

如果主账号登录后，发现已经可以直接创建环境，说明角色授权已完成，无需再次操作，可跳过此步骤。

步骤2：主账号为子账号关联策略

1. 用主账号登录 [访问管理控制台](#) > [用户](#) > [用户列表](#)，找到需要授权的字账号，单击对应的授权。



2. 在弹出的关联策略弹窗中，搜索以下两个策略：QcloudCamReadOnlyAccess，QcloudAccessForTCBRole，选中后单击确定。

关联策略
✕

选择策略 (共 3 条)

策略名	策略类型
<input checked="" type="checkbox"/> QcloudAccessForTCBRole 云开发 (TCB) 对云资源的访问权限	预设策略
<input type="checkbox"/> QcloudAccessForTCBRoleInAccessCloudB... 该策略供云开发 (TCB) 服务角色 (TCB_Q...	预设策略
<input type="checkbox"/> QcloudAccessForTCBRoleInCBMClient 该策略供云开发 (TCB) 服务角色(TCB_Qcs...	预设策略

已选择 2 条

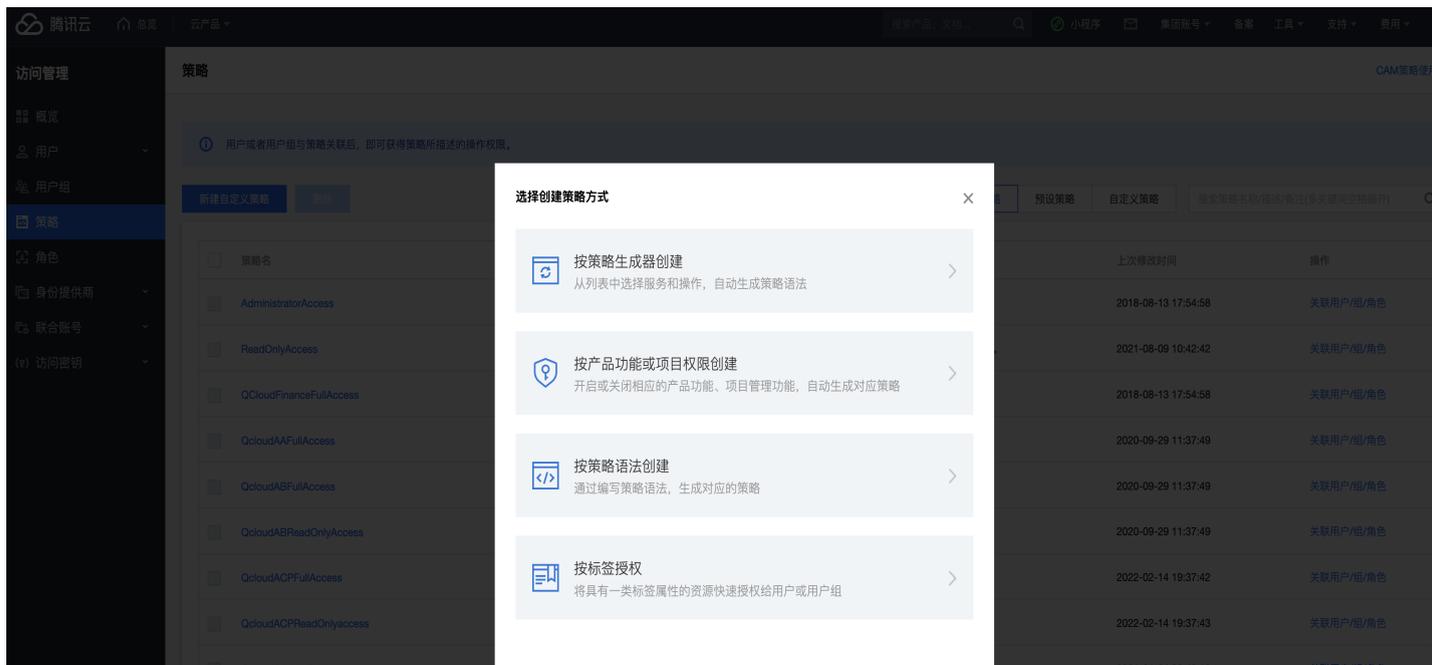
策略名	策略类型
QcloudCamReadOnlyAccess 用户与权限 (CAM) 只读访问权限	预设策略 ✕
QcloudAccessForTCBRole 云开发 (TCB) 对云资源的访问权限	预设策略 ✕

支持按住 shift 键进行多选

确定
取消

步骤3: 创建自定义策略

用主账号登录 [访问管理控制台](#) > [策略](#)，单击新建自定义策略，选择按策略生成器生成。



按以下字段说明编辑策略：

- 效果（Effect）：允许表示授予子账号可以做某操作，拒绝表示禁止子账号做某操作。
- 服务（Service）：选择云托管 **CloudBase Run**。
- 操作（Action）：可选全部或只允许或拒绝子账号能对云托管做指定的操作。可进一步细分为读操作或写操作。
- 资源（Resource）：
 - 选择**全部资源**，则子账号可以操作所有的云托管独立控制台环境（不包含云开发中云托管）。
 - 选择**特定资源**，则子账号只能操作指定的云托管独立控制台环境，不能操作其他环境。资源六段式：在**特定资源**中，需要通过资源六段式指定策略生效范围。
 - 服务：默认填写“tibr”。
 - 地域：请按实际地域选择。（目前云托管仅支持北京/上海/广州）。
 - 账户：默认填写“uin/\${主账号 ID}”。
 - 资源前缀：默认填写“env”。
 - 资源：请按实际环境 ID，填写“\${envId}”。

下面是一个配置好了的策略示例。配置效果：被这个策略关联的用户、用户组、角色，只能操作账号下“test-env”这个环境，能对这个环境做全部操作。

访问管理

← 按策略生成器创建

1 编辑策略 > 2 关联用户/用户组/角色

可视化策略生成器 JSON

▼ 云托管 CloudBase Run (全部操作)

效果 (Effect) * 允许 拒绝

服务 (Service) * 云托管 CloudBase Run (tcsr)

操作 (Action) * 全部操作 (*)

资源 (Resource) * 全部资源 特定资源

收起

不拆分资源级和操作级接口 ⓘ

env qcs::tcsr::uin/100007::env/test-env 编辑 删除 此类型任意资源

[添加资源六段式 来限制访问](#)

[添加自定义资源六段式 来限制访问](#)

条件 (Condition) 来源 IP ⓘ

[添加其他条件](#)

+ 添加权限

下一步 字符数: 161 (最多6144)

使用 JSON 方式编辑策略，输入以下内容也可得到示例效果：

```

{
  "version": "2.0",
  "statement": [
    {
      "effect": "allow",
      "action": [
        "tcsr:*"
      ],
      "resource": [
        "qcs::tcsr::uin/100007xxxxxx:env/test-env"
      ]
    }
  ]
}
    
```

```
    ]  
  }  
]  
}
```

监控告警设置

最近更新时间：2024-08-05 15:23:11

⚠ 注意：

本篇文章仅适用于云托管新版控制台。

云托管独立控制台中，已经提供了一系列简单的监控图表。如果您需要更进一步详细的监控视图及告警能力，需要配合使用腾讯云可观测平台产品。

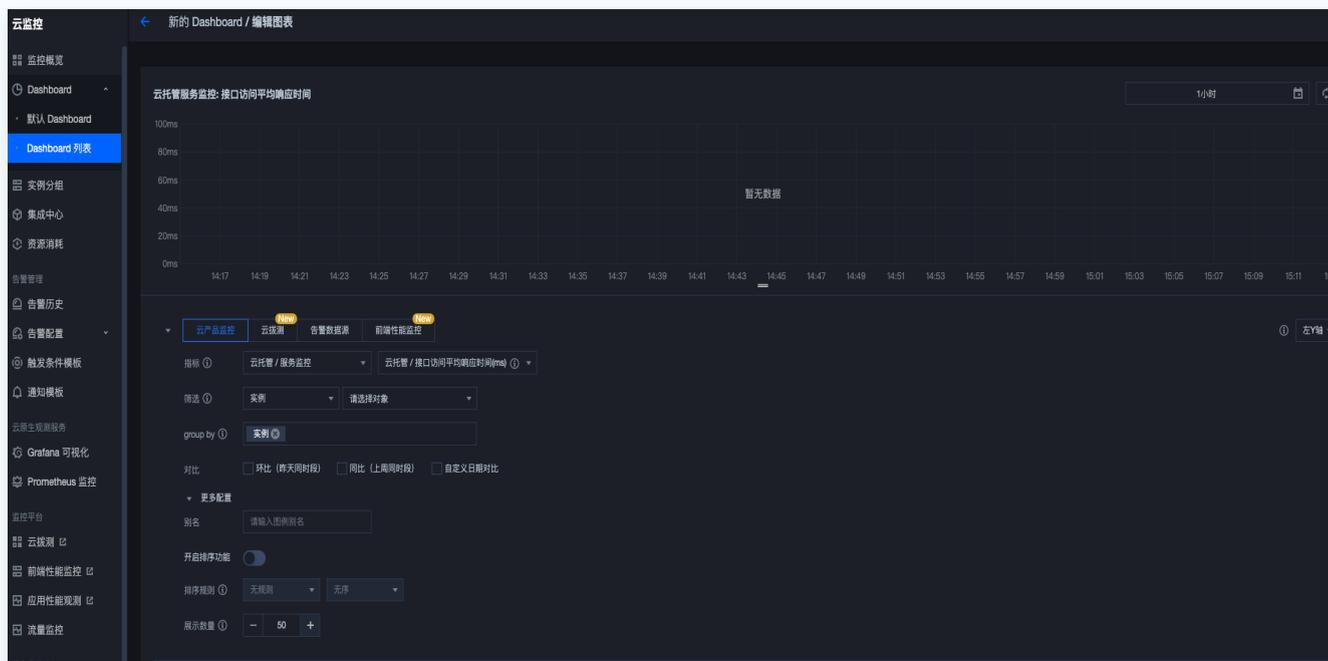
操作步骤

步骤1：新建监控 Dashboard

登录 [腾讯云可观测平台](#)，单击新建 Dashboard > 新建图表，选择云产品监控 > 云托管 > 服务监控。

ⓘ 说明：

关于什么是 Dashboard 图表，请参见 [腾讯云可观测平台文档](#)。



步骤2：新建云托管告警策略

在 Dashboard > 编辑图表中，监控类别选择云产品监控，策略类型选择云托管 > 服务监控。告警对象可以按实际的地域 > 环境 > 服务选择。



步骤3：配置告警触发条件和告警通知

请参见 [腾讯云可观测平台文档-告警服务](#)。

字段说明

指标	说明
调用量	单位时间内服务收到的请求次数总和（包括通过默认公网域名、自定义域名）。
平均响应时间	单位时间内服务收到的所有请求的响应时间取平均值。
HTTP 错误	单位时间内请求服务失败返回 HTTP 错误的次数。
CCU 使用量	单位时间内服务所有版本容器算力资源消耗之和，单位为（个 x 小时）。规格1C2G运行一小时消耗1CCU，依此类推。

平均实例个数	服务所有版本的实时实例个数之和，在单位时间内取平均值。
异常实例个数	服务所有版本的实时实例状态不正常个数之和，在单位时间内取平均值。

云开发中云托管配置指南

服务配置说明

最近更新时间：2024-06-05 17:19:31

⚠ 注意：

本篇文章仅适用于云开发中的云托管。

服务名称

环境下服务的唯一标识。创建成功后**不支持修改服务名称**。如果您有多个环境，不同环境之间服务可以重名。（例如：开发环境和测试环境可以同时存在服务 "testservice"）

备注信息

展示在服务列表页面，仅您自己在控制台可见，不对部署产生任何影响。

所在私有网络/云托管网络/子网

背景知识

关于什么是私有网络（VPC）及子网，请参见 [私有网络](#) 文档。单击服务所在私有网络的名称，可以跳转到 [私有网络控制台](#) 查看私有网络的详情，以及该私有网络内您还有哪些其他云资源，可以与此私有网络内的云托管服务配合使用。

使用限制

- 如果您的服务需要访问腾讯云数据库或其他云服务，您可以将这些外部依赖部署于同个私有网络中。
- 不支持更换服务所在私有网络**，您可以在正确的私有网络下重新创建服务后，再删除当前服务。

用法推荐

- 新建本环境内首个服务，且无特别的网络要求，请选择**系统创建**，云托管会自动为您创建并配置好一个全新的VPC，同时**自动创建一个全新的微型 NAT 网关**，保证您新建的服务具备访问公网的能力。
- 新建本环境内的第 2 ~ 第 N 个服务，且无特别的网络要求，推荐选择和前一个服务相同的私有网络。
- 如果有明确的网络要求，请选择**已有私有网络**，指定符合您要求的 VPC 及子网。

📌 说明：

若您选择的 VPC 不是由云托管创建的，需要您自行在此 VPC 中配置 NAT 网关，否则您的服务将无法访问公网。

NAT 网关

背景知识

关于什么是 NAT 网关及如何配置公网访问，请参见 [NAT 网关](#) 文档。

使用限制

- 云托管自动创建的 NAT 网关（含“系统创建”标签），为云托管定制的微小型网关，最大并发数 10 万，最高带宽 10Mbps。如果不满足您的使用需求，请前往 [NAT 网关控制台](#) 新增并配置符合您要求的 NAT 网关。
- 非云托管创建的 NAT 网关，在云托管控制台无法展示 IP，费用由 NAT 网关侧进行结算。请前往 [NAT 网关控制台](#) 查看并管理。

用法推荐

NAT 网关 IP，将作为您的服务访问公网的出口 IP，也可用作域名备案。但无法通过这个 IP 访问到您的服务。

镜像仓库

背景知识

云托管使用的是容器镜像服务（TCR）个人版，没有费用。您可以在服务详情页面中的**镜像**选项卡中，快速查看当前服务所绑定的镜像仓库中所有镜像和它们的使用情况。更多关于腾讯云镜像仓库的介绍，请参见 [容器镜像服务](#) 文档。

使用限制

- 服务在创建时会绑定一个唯一的腾讯云镜像仓库，管理该服务下的所有版本的相关镜像。
- 不支持同一服务绑定多个镜像仓库，使得不同版本的镜像分散在多个不同的镜像仓库里。
- 所有由云托管创建的镜像仓库都会统一存放于以“tcb”开头的同一个命名空间之下，并以环境 `id_` 服务名的格式命名。您可以在 [容器镜像服务控制台](#) 上看到由云托管创建的这个镜像仓库并对它进行管理。
- 暂不支持服务绑定公有仓库。

用法推荐

使用系统默认仓库（推荐）：创建服务时，云托管将自动为您创建一个与服务同名的私有仓库并与您的服务进行绑定。绑定已有腾讯云镜像仓库适用以下场景：

跨场景复用

在云托管的场景之外，您已经使用了腾讯云的镜像仓库，并希望里面的镜像继续用于云托管的某个服务。此时，在创建服务时，可以选择“绑定已有腾讯云镜像仓库”，指定任意您账户下的私人镜像仓库。

请注意，若如此做，则您通过云托管产生的镜像（手动上传或自动构建推送），可能也会影响您在其他场景下的镜像使用。

多个服务共用同一套代码/镜像

如果同一个服务需要同时部署 N ($N \geq 2$) 套（在同一环境部署多套请用不同服务名加以区分，分别部署在不同环境服务可以重名），则在创建第 2 - N 个服务时，可以选择“绑定已有腾讯云镜像仓库”，并指定第一个服务所绑定的镜像仓库，实现镜像在多个服务间的复用。

两个服务都绑定到了同一个镜像仓库，仓库中所有的镜像可以同时用于两个服务，且这两个服务未来由云托管自动构建的镜像都会被推送到这同一个镜像仓库。

特殊情况

如果您曾经创建了一个服务并选择使用系统默认仓库，后来又删除了这个服务，但保留了这个同名的镜像仓库。则当您在同一环境再次用同样的名字创建服务时，选择使用系统默认仓库会自动绑定到原有的这个同名镜像仓库，而不会再次创建一个全新的镜像仓库。

此时，若您愿意继续使用之前同名服务遗留的镜像，可直接开始使用；若您不再需要遗留的镜像，请您到服务详情页面的镜像选项卡中，或到 [容器镜像服务控制台](#) 手动删除不需要的镜像。

日志管理/日志设置

背景知识

如何在云开发控制台内查询服务日志，请参见 [服务日志查询](#)。关于腾讯云 Elasticsearch service 的介绍，请参见 [Elasticsearch service](#) 文档。

使用限制

- 服务可以选择将日志输入到云开发日志管理模块中，或输出到指定的 ES 中。创建服务成功后不支持修改日志选项。
- 选择自定义 Elasticsearch，则在云开发或云托管控制台上无法查看您的服务日志，请前往 [Elasticsearch service 控制台](#) 查看。

服务域名

- 默认公网域名：公网任意来源可以用来对服务进行 HTTPS 调用，会产生相应的公网流量费用。
- 允许公网通过服务域名访问服务：关闭后，通过默认公网域名/自定义公网域名访问服务将报错。

HTTP 访问服务配置

在服务域名之外，可选择性启用 HTTP 访问服务。

背景知识

关于云开发 HTTP 访问服务的介绍，请参见 [HTTP 访问服务文档](#)。

使用限制

- 是否启用、如何配置 HTTP 访问服务，均不影响对服务域名的使用。
- 关闭允许公网通过服务域名访问服务开关，不影响公网通过在 HTTP 访问服务中配置的地址访问服务。
- 公网通过在 HTTP 访问服务中配置的地址访问服务，会产生相应的公网流量费用。

版本配置说明

最近更新时间：2025-02-26 15:11:02

⚠ 注意：

本篇文章仅适用于云开发中的云托管。

基础配置说明

上传方式

支持本地代码、代码库拉取和镜像拉取三种方式。

方式1：本地代码

上传代码工程文件夹或 zip 压缩包，系统解析成功后先将代码构建为镜像、将镜像推送到服务绑定的镜像仓库储存，然后基于该镜像进行部署。

此方式需要您自行编写 Dockerfile 并包含在代码文件夹或 zip 压缩包中。

📌 说明：

我们提供 [Java](#)、[PHP](#)、[Node.js](#)、[Golang](#)、[Python](#) 语言的示例代码包下载用于 Demo。除此之外的任意开发语言也都可支持。如需更多语言的 Demo 代码包资源，请 [联系我们](#)。

方式2：代码库拉取

获得您的授权后，从您的 GitHub、GitLab 或码云 Gitee 仓库拉取代码后，先将代码构建为镜像、将镜像推送到服务绑定的镜像仓库储存，然后基于该镜像进行部署。

此方式需要您自行编写 Dockerfile 并包含在代码库中。

📌 说明：

首次使用需要您先进行授权。授权成功后，当前环境任意服务新建版本都不再需要再次授权。

方式3：镜像拉取

无需系统提供从代码构建镜像的服务，由您自行采用任意方式构建好镜像后，手动将镜像上传至服务所绑定的镜像仓库中，然后系统基于您选定的镜像进行部署。

📌 说明：

- 如何上传镜像，请查看本文档“相关操作”中的“手动上传镜像”部分。
- 使用 Demo 镜像可快速体验部署过程。选择 Demo 镜像后无需修改任何配置，直接开始部署。

监听端口

应用监听端口。默认为 80。

流量策略

部署完成后，如何开启流量。目前有如下两种选择：

- 部署完成后保持流量为 0 稍后再手动调整流量：希望部署完后先手动验证再开流量，或只希望引入部分流量进行灰度升级。
- 部署完成后自动开启 100%流量：首次部署完成后立刻开流量，或升级时进行全量升级。

备注信息

选填，会展示在版本列表页面中，对部署不产生任何影响。

自动扩缩容配置说明

副本模式

支持低成本或高可用两种模式。

模式1：低成本

适合对成本敏感，对冷启动相对不敏感的业务。

- 副本个数最小值为 0，当没有流量打到版本上时，版本将缩容到 0，不保留任何实例，不产生任何费用。
- 连续半个小时无流量才将实例缩容到 0（避免流量偶然波动带来的误判）。再次冷启动时，可能有 30 秒服务延迟。
- 默认推荐规格为最小规格 0.25 核 0.5G，**单价最低**。您也可按需调整规格。
- 虽然没有业务流量，部署过程中仍然会先产生一个实例，部署完成后再缩容到 0，因此**部署过程本身会产生一定的资源消耗**。
- 是否有业务流量，是通过是否有通过域名访问来判断的。如果通过内网 IP 调用（不推荐），或者服务跑的是定时任务没有外部调用，也会被认为无流量触发缩容。以上类型业务，请采用高可用模式。

ⓘ 说明：

将版本流量百分比设置为 0 并不是触发缩容到 0 的必要条件。低成本模式下，版本流量百分比设置为 0，一定会触发缩容到 0。但如果版本流量百分比不为 0，需要连续半小时的观测期，期间版本没有产生真实业务流量，才会触发缩容到 0。

模式2：高可用

适合对成本相对不敏感，对服务常驻有诉求，或无法接受冷启动的业务。

- 副本个数最小值不能为 0，可以设置为 1 ~ 50 间的任意整数。即便没有流量打到版本上，仍会保持最小个数的实例数，也会**持续产生费用**。

- 部署过程中即开始按最小副本个数和规格产生实例，部署过程本身会产生一定的资源消耗。

规格

指集群中每个容器实例的配置。自动扩容时，新创建的实例将使用这个规格。同一个版本下所有容器实例规格都会保持一致。

副本个数

指当前版本在自动扩缩容时可达到的最大实例数及最小实例数。最小值下限为 0，最大值上限为 50。

- 如需将最小值修改为 0，请先切换副本模式至“低成本”。
- 如需将最小值修改为大于 0 的整数，请先切换副本模式至“高可用”。

扩缩容条件

当达到某个条件时，云托管会自动创建/删除实例，然后检测是否再次达到条件，如果满足条件则继续扩缩容，如此反复直至实例数量达到副本个数的最小值/最大值，或不再满足扩缩容条件时停止自动扩缩容。

目前仅支持 CPU 使用率作为扩缩容条件，更多扩缩容指标敬请期待。

! 说明：

- 自动扩容到副本个数最大值后若仍不足以承载业务流量，即便再次达到扩缩容条件，也不会继续创建新的实例，可能导致您的业务受影响，请您评估好业务指标后合理设定副本个数最大值。
- 若希望提升最大值限额（大于 50），请提交工单联系我们单独处理。

高级配置说明

高级配置可不做任何修改直接使用默认值开始部署。

InitialDelaySeconds

实例创建完成后，等待一定的时间后开始进行健康检查。若健康检查失败，将重试 3 次，依然失败则判定服务版本异常。请将此值设定为大于应用启动的时间，否则版本可能持续处于异常状态。默认值为 2 秒。

日志采集路径

可设置目录或文件，支持设置多个路径。采集到的日志可以在 [日志管理](#) 中查看。支持标准输出（stdout）以及通配路径（例：/logs/*），使用 `,`（半角逗号）分割，留空将采集标准输出。

环境变量

用户所需的环境变量，直接传入容器中。以 key value 的形式可配置多个。

流量配置说明

最近更新时间：2024-06-05 17:19:31

⚠ 注意：

本篇文章只适用于云开发中的云托管。

按百分比（随机）分配流量

1. 单击**流量配置**。
2. 模式采用默认**按百分比**。
3. 选中刚部署完的最新版本和当前运行中的旧版本，并为每个版本指定流量百分比。所有百分比加起来需要等于100%或0%，否则无法完成。未在流量配置窗口中添加的版本，将统一将流量百分比变为0%。
4. 单击**完成**，系统立刻开始按百分比随机分配流量，将部分流量从旧版本引向新版本。

流量配置

操作提示

1. 所有版本流量百分比之和必须等于100%或0%。
2. 未设定流量百分比的版本，将统一将流量百分比变为0%。

模式 按百分比 按 URL 参数

mytest-012	50	%	✕
mytest-001	50	%	✕

[新建](#)

成功后，可以看到对应版本的流量百分比已经变化。

版本	状态	流量	备注	上传方式	创建/更新时间	操作
mytest-012	正常	50%	-	镜像拉取	2020-08-28 16:48:55 2020-09-20 16:28:20	删除 更多
mytest-009	构建失败	0%	-	本地代码	2020-07-10 17:23:10 2020-07-10 17:23:23	重新构建 删除 更多
mytest-001	正常	50%	-	镜像拉取	2020-07-07 15:42:04 2020-09-06 18:37:01	删除 更多

共 3 条

20 条 / 页

按 URL 参数（定向）分配流量

1. 单击**流量配置**。
2. 模式选择**按 URL 参数**。
3. 添加部署完的最新版本，设定 URL 参数 key 和 value（支持 Glob 表达式）。符合条件的请求将路由到新版本。
4. 再添加当前运行中的旧版本和对应的 URL 参数。
5. 将旧版本设置为默认版本。
6. 在**匹配结果预测**中输入 URL 参数模拟请求，单击**预测**查看对应版本，调试配置。
7. 单击**完成**，系统立刻开始将部分流量从旧版本引向新版本。

流量配置 ✕

操作提示

1. 请配置至少一个优先级条件和一个默认版本。
2. 存在多个优先级条件时，将按优先顺序进行完全匹配。若同时符合优先级1和优先级2的条件，则该请求将被路由到优先级1中的版本。不满足任何一个优先级条件的流量全部流入默认版本。
3. 配置完优先级条件和默认版本后，您可以在“匹配结果预测”中输入参数补全URL，预测此请求将被路由到哪个版本，用以调试您的配置。

模式 按百分比 按 URL 参数

优先级1 ✕

优先级2 ✕

添加

默认版本 ?

匹配结果预测 http://test-2-318335.service.tcloudbase.com/cls-pv8fcqrj-mytest?

预测

对应版本: mytest-012

完成
关闭

成功后，可以看到对应版本的流量条件已经变化。

版本	状态	流量	备注	上传方式	创建/更新时间	操作
mytest-012	正常	version=new	-	镜像拉取	2020-08-28 16:48:55 2020-09-20 16:28:20	删除 更多
mytest-009	构建失败	-	-	本地代码	2020-07-10 17:23:10 2020-07-10 17:23:23	重新构建 删除 更多
mytest-001	正常	version=old 默认版本	-	镜像拉取	2020-07-07 15:42:04 2020-09-06 18:37:01	删除 更多

共 3 条
20 条 / 页