

# Tencent Kubernetes Engine Services Product Introduction



## Copyright Notice

©2013-2018 Tencent Cloud. All rights reserved.

Copyright in this document is exclusively owned by Tencent Cloud. You must not reproduce, modify, copy or distribute in any way, in whole or in part, the contents of this document without Tencent Cloud's the prior written consent.

## Trademark Notice



All trademarks associated with Tencent Cloud and its services are owned by Tencent Cloud Computing (Beijing) Company Limited and its affiliated companies. Trademarks of third parties referred to in this document are owned by their respective proprietors.

## Service Statement

This document is intended to provide users with general information about Tencent Cloud's products and services only and does not form part of Tencent Cloud's terms and conditions. Tencent Cloud's products or services are subject to change. Specific products and services and the standards applicable to them are exclusively provided for in Tencent Cloud's applicable terms and conditions.

# Contents

## Services

- Basic Operations of Services

- Setting Limit on Service Resources

- Setting Access Mode of Services

# Services

## Basic Operations of Services

Last updated : 2018-07-05 15:56:53

### Creating Service

1. Log in to the [Tencent Cloud TKE console](#).
2. Click **Service** in the left navigation bar, and click **New** in the service list page.



3. Configure basic service information.

- **Service Name:** The name of the service to be created, which has a length limited to 63 characters comprised of lowercase letters, numbers and "-". It starts with a lowercase letter and ends with a lowercase letter or a number.
- **Region:** Select the region where the service is deployed.
- **Cluster:** Select a cluster to run your service. You need to select a running cluster with available CVMs in it.
- **Service description:** Information of created service. This information is displayed in **Service Information** page.

[< 返回](#) | **新建服务**

---

服务名称   
最长63个字符，只能包含小写字母、数字及分隔符("-")，且必须以小写字母开头，数字或小写字母结尾

所在地域 广州 上海 北京 新加坡 香港

运行集群    
如现有的集群不合适，您可以去控制台 [新建集群](#) 或 [新建Namespace](#)

服务描述

#### 4. Configure volume.

Click **Add Volume** when you specify a specific path to which a container is mounted.

**Note:**

If no source path is specified, a temporary path is assigned by default.

- **Type:** Four types of volumes are supported: local disk, cloud disk, NFS disk, and configuration file. For more information, please see [How to Use TKE Volume](#).
- **Name:** Volume name.
- **Path:** Specify the path to which a container is mounted.

#### 5. Configure a container.

- **Name:** The name of the container to be created, with a length limited to 63 characters.
- **Image:** Click **Select Image** to create a service under My Images, My Favorites, TencentHub Image, DockerHub Image or other images.
- **Tag:** The default image tag for TKE. If you need to use a different image tag, click tag display window to select one.

### 运行容器

名称   
最长63个字符，只能包含小写字母、数字及分隔符("-")，且不能以分隔符开头或结尾

镜像  [选择镜像](#)

镜像版本 (Tag)

资源限制

CPU限制				内存限制			
request	0.25	-	limit 0.5 核	request	256	-	limit 1024 MIB

Request用于预分配资源,当集群中的节点没有request所要求的资源数量时,容器会创建失败。  
Limit用于设置容器使用资源的最大上限,避免异常情况下节点资源消耗过多。

环境变量 ⓘ [新增变量](#) [从配置项导入](#)  
变量名只能包含大小写字母、数字及下划线，并且不能以数字开头

[显示高级设置](#)

注意：服务创建完成后，容器的配置信息可以通过更新服务的方式进行修改

添加容器

## 6. Other settings

- **Number of pods:** A pod consists of one or more relevant containers. You can specify the number of pods by clicking + or -.
- **Service access method:** The method for accessing a service determines the network attribute of this service. Different access methods offer different network capabilities. For more information about the

four access methods, please see [Configuration of Service Access Methods](#).

实例数量

服务访问方式 <sup>①</sup>  提供公网访问  仅在集群内访问  VPC内网访问  不启用 (不支持Ingress) [如何选择](#)

将提供一个可以从Internet访问入口,支持TCP/UDP协议,如web前台类服务可以选择公网访问。

如您需要公网通过HTTP/HTTPS协议或根据URL转发,您可以在Ingress页面使用Ingress进行路由转发, [查看详情](#)

端口映射

协议 <sup>①</sup>	容器端口	服务端口 <sup>①</sup>
TCP <input type="button" value="v"/>	<input type="text" value="容器内应用程序监听的端口"/>	<input type="text" value="建议与容器端口一致"/> <input type="button" value="x"/>

[添加端口映射](#)

7. Click **Create Service** to complete the creation process.

## Updating the Number of Pods

1. Click **Services** in the left navigation bar of TKE console to enter the service list, and click **Update Number of Pods**.

容器服务

服务 广州 上海 北京 新加坡 香港

所属集群 dls-kza53cvy (t) 所属集群空间 default 服务操作文档

+ 新建

名称	监控	日志	状态	运行/预...	IP地址	负载均衡	标签(l...)	创建时间	操作
nginx			运行中	1/1个	111.230.16... 172.16.255...	lb-gkuy6...	qcloud-app...	2017-12-22 10...	<span style="border: 2px solid red; padding: 2px;">更新实例数量</span> <span>更新服务</span> <span>更多</span>

共1项
每页显示行 20

⏪ ⏩ 1/1 ⏪ ⏩

2. You can specify the number of pods by clicking + or -. Click **OK** after the configuration is completed.

## Updating Service

1. Click **Services** in the left navigation bar of TKE console to enter the service list, and click **Update Service**.



The screenshot shows the Tencent Cloud TKE console interface. On the left is a dark navigation sidebar with '容器服务' (Container Service) selected. The main area displays a table of services. The table has columns for Name, Monitoring, Logs, Status, Running/Preparation, IP Address, Load Balancing, Labels, Creation Time, and Actions. One service named 'nginx' is listed with a status of '运行中' (Running). In the '操作' (Actions) column for this service, the '更新服务' (Update Service) button is highlighted with a red rectangular box.

2. Click **Start Update**.

Two updating methods are available.

- **Rolling update**: Update pods one by one, which allows you to update the service without interrupting the business.
- **Quick update**: Directly close all current pods, and launch the same number of new pods.

## Redeployment

Redeployment is to redeploy the containers under a service and pull a new image.

1. Click **Services** in the left navigation bar of TKE console to enter the service list, and click **More** -> **Redeploy**.



The screenshot shows the Tencent Cloud TKE console interface. On the left is a navigation sidebar with '容器服务' (Container Service) selected. The main area displays a table of services. The first service is 'nginx', which is in a '运行中' (Running) state. The '操作' (Actions) column for this service has a '更多' (More) dropdown menu open, and the '重新部署' (Redeploy) option is highlighted with a red box. Other options in the menu include '更新实例数量' (Update Instance Count), '更新服务' (Update Service), '查看Yaml' (View Yaml), and '删除' (Delete).

名称	监控	日志	状态	运行/预...	IP地址	负载均衡	标签(L...	创建时间	操作
nginx			运行中	1/1个	111.230.16... 172.16.255...	lb-gkuy6...	qcloud-app...	2017-12-22 10...	<a href="#">更新实例数量</a> <a href="#">更新服务</a> <a href="#">更多</a> <a href="#">重新部署</a> <a href="#">查看Yaml</a> <a href="#">删除</a>

2. Click **OK**.

## Deleting Service

1. Click **Services** in the left navigation bar of TKE console to enter the service list, and click **More** -> **Delete**.



The screenshot shows the Tencent Cloud Container Service console. On the left is a navigation sidebar with options like '概览', '应用中心', '集群', '服务', 'Ingress', etc. The main area displays a table of services. The table has columns for '名称' (Name), '监控' (Monitoring), '日志' (Logs), '状态' (Status), '运行/预...' (Running/Pre-...), 'IP地址' (IP Address), '负载均衡' (Load Balancing), '标签' (Tags), '创建时间' (Creation Time), and '操作' (Actions). One service named 'nginx' is listed with status '运行中' (Running). A context menu is open over the 'nginx' row, showing options: '更新实例数量' (Update Instance Count), '更新服务' (Update Service), '更多' (More), '重新部署' (Redeploy), '查看Yaml' (View Yaml), and '删除' (Delete). The '删除' option is highlighted with a red box.

名称	监控	日志	状态	运行/预...	IP地址	负载均衡	标签	创建时间	操作
nginx			运行中	1/1个	111.230.16... 172.16.255...	lb-gkuy6...	qcloud-app...	2017-12-22 10...	更新实例数量 更新服务 更多 重新部署 查看Yaml 删除

2. Click **OK**.



The screenshot shows a confirmation dialog box titled '删除服务' (Delete Service). The dialog asks: '您确定要删除服务"nginx"吗?' (Are you sure you want to delete the service 'nginx'?). Below the question, it says: '该服务下所有实例和外网负载均衡将一并销毁，请提前备份好数据。' (All instances and public network load balancers under this service will be destroyed together. Please back up your data in advance.) At the bottom of the dialog are two buttons: '确定' (OK) and '取消' (Cancel). The '确定' button is highlighted in blue.

**Note:**

Deleting the service will delete all the pods and public network load balancers under the service. Back up your data in advance.

# Setting Limit on Service Resources

Last updated : 2018-07-05 15:55:16

## Resource Limits on Container Services

### About "Request" and "Limit"

**request** : The minimum amount of resources used by the container, a criterion for resources allocation when you schedule the container. The container is scheduled to the node only when the amount of resource that can be assigned on the node  $\geq$  the number of container resources requests. Parameter Request does not limit the maximum available resources for the container.

**limit** : The maximum amount of resources used by the container. A value of 0 indicates no upper limit to the amount of resources.

For more information on **limit** and **request**, click [here](#)

### CPU Limits

You can set Request and Limit for the amount of CPU, which is measured in cpu units (U) and expressed with decimal points.

1. CPU Request is used as a criterion for users to assign CPU resources for the container on the node during creation. It is called the Assigned CPU Resources.
2. CPU limit specifies the maximum amount of CPU to reserve for a container. A value of 0 indicates no upper limit to the amount of CPU (CPU Limit  $\geq$  CPU Request).

### Memory Limits

You can only set the maximum amount of memory available for the container. Memory is measured in MiB, and expressed with decimal points.

1. Memory can't be scaled. When memory used by the container exceeds Memory Request, the container may be killed. Therefore, to ensure normal operation of the container, Request should be equal to Limit.
  - i. Memory Request (=Limit) is used as a criterion for users to assign memory resources for the container on the node during creation. It is called the Assigned Memory Resources.

### CPU Usage Vs. CPU Utilization

1. CPU usage indicates the number of physical CPUs used by the container. CPU usage is used to determine whether CPU resource exceeds the Request and the Limit.

2. CPU utilization is the ratio of CPU usage to number of CPU single cores (or number of CPU cores on the node)

### Example

A simple example is given here to illustrate the role of Request and Limit. A 4U4G node is provided in the trial cluster. Two Pods (1, 2) are deployed, each with resources configuration of (CPU Request, CPU Limit, Memory Request, Memory Limit) = (1, 2U, 1G, 1G).

(1.0 GB = 1000 MiB)

The usage of CPU and memory on the node is shown as below:



Assigned CPU resources: 1U (to Pod 1) + 1U (to Pod 2) = 2U; unassigned CPU resources: 2U

Assigned memory resources: 1G (to Pod 1) + 1G (to Pod 2) = 2G; unassigned memory resources: 2G

In such case, one Pod with (CPU Request, Memory Request) = (2U, 2G), or two Pods with (CPU Request, Memory Request) = (1U, 2G) can be deployed on the node.

For Pod (1, 2), (CPU Limit, Memory Limit) = (2U, 1G). That is, the maximum amount of CPU available for a Pod is 2U when the resource is idle.

# Setting Access Mode of Services

Last updated : 2018-07-05 15:56:07

The method for accessing a service determines the network attribute of this service. Different access methods offer different network capabilities. Tencent Cloud TKE provides four service access methods: Access from Public Network, Access Within Cluster Only, Access via Private Network in VPC and Not Enabled.

## Access from Public Network

For services accessed from public network, an entry (the public network load balancer) for Internet access is provided. The service accessed using this method can be accessed directly from the public network. This method is applicable to Web frontend services, such as wordpress frontend service.

For example, if we create a wordpress service that can be accessed from public network, the service access method is set to **Access From Public Network**. The created service can be accessed directly through **LB IP + Service Port**.

For more information on how to create a wordpress service, please see [WordPress with Single Pod](#).

## Access Within Cluster Only

For services accessed within a cluster, an entry (service IP) of being accessed by other services or containers within the cluster is provided. This method is applicable to MySQL and other database services to ensure the isolation among service networks.

For example, if we create a MySQL service that can be accessed within a cluster only, the service access method is set to **Access Within Cluster Only**. The created service can be accessed directly through **Service IP/Name + Service Port**.

## Access via Private Network in VPC

For services accessed via the private network in a VPC, an entry (private network load balancer) of being accessed by other resources under the VPC in which the cluster resides is provided. This method is

applicable to services that need to be accessed by other clusters or CVMs under the same VPC. For example, if we create a MySQL service that can be accessed via the private network in a VPC, the service access method is set to **Access via Private Network in VPC**. The created service can be accessed directly through **Private network LB IP + Service Port**.

## Not Enabled

If "Not Enabled" is selected for service access, no entry is provided for accessing from frontend service to the container. This can be used to discover or easily enable multiple container pods with custom service.

## More

In addition to the above four methods, you can also configure a layer-7 load balancer (HTTP/HTTPS) to forward to the service. For more information, please see [Ingress Forwarding Configuration](#).

For more information on how to access within a cluster, please see [Multiple Service Access Methods in Kubenerters and Configuration of Security Group in Tencent Cloud](#).