

Tencent Kubernetes Engine

Purchase Guide

Product Documentation



Copyright Notice

©2013-2019 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

Purchase Guide

Purchasing Container Clusters

- Regions and Availability Zones

- Quota Limits for Cluster Purchase

- Network Settings for Containers and Nodes

- Container Node Disk Settings

- Notes on the Public IP of a TKE Node

- TKE Security Group Settings

- Project of New Resources

Purchase Elastic Clusters

- Billing Overview

- Product Pricing

- Purchase Limits

- Resource specification

Billing Description

Purchase Channels

Purchase Guide

Purchasing Container Clusters

Regions and Availability Zones

Last updated : 2020-10-26 16:18:02

Regions

Introduction

A region is the physical location of an IDC. In Tencent Cloud, different regions are fully isolated, ensuring maximum stability and fault tolerance among regions. To reduce access latency and increase download speeds, we recommend you select the region nearest to your customers.

You can view the following table or use the [DescribeRegions](#) API to get a complete list of regions.

Characteristics

- The networks of different regions are fully isolated. Tencent Cloud services in different regions **cannot communicate via a private network by default**.
- Tencent Cloud services across regions can communicate with each other through [public IPs](#) over the Internet, while those in different VPCs can communicate with each other through [CCN](#) that is faster and steadier.
- [Cloud Load Balancer](#) currently supports intra-region traffic forwarding, and binding CVMs in the same region. If the [cross-region binding](#) feature is enabled, then binding CVMs to Cloud Load Balancer across regions is supported.

Availability Zone

Introduction

An availability zone is a physical IDC of Tencent Cloud with independent power supply and network in a single region. It can ensure business stability, as failures in one AZ are isolated without affecting other AZs in the same region. By starting an instance in an independent availability zone, users can protect their applications from being affected by a single point of failure.

You can view the following table or use the [DescribeZones](#) API to get a complete list of availability zones.

Characteristics

Tencent Cloud services in the same VPC are interconnected through the private network, meaning they can communicate using [private IPs](#), even if they are in different availability zones.

Note :

Private network interconnection refers to the interconnection of resources under the same account. Resources under different accounts are completely isolated on the private network.

China

Region	AZ
South China (Guangzhou) ap-guangzhou	Guangzhou Zone 1 (sold out) ap-guangzhou-1
	Guangzhou Zone 2 (sold out) ap-guangzhou-2
	Guangzhou Zone 3 ap-guangzhou-3
	Guangzhou Zone 4 ap-guangzhou-4
	Guangzhou Zone 6 ap-guangzhou-6
East China (Shanghai) ap-shanghai	Shanghai Zone 1 ap-shanghai-1
	Shanghai Zone 2 ap-shanghai-2
	Shanghai Zone 3 ap-shanghai-3
	Shanghai Zone 4 ap-shanghai-4
	Shanghai Zone 5 ap-shanghai-5
East China (Nanjing)	Nanjing Zone 1

ap-nanjing	ap-nanjing-1
	Nanjing Zone 2 ap-nanjing-2
North China (Beijing) ap-beijing	Beijing Zone 1 ap-beijing-1
	Beijing Zone 2 ap-beijing-2
	Beijing Zone 3 ap-beijing-3
	Beijing Zone 4 ap-beijing-4
	Beijing Zone 5 ap-beijing-5
Southwest China (Chengdu) ap-chengdu	Chengdu Zone 1 ap-chengdu-1
	Chengdu Zone 2 ap-chengdu-2
Southwest China (Chongqing) ap-chongqing	Chongqing Zone 1 ap-chongqing-1
Hong Kong/Macao/Taiwan (Hong Kong, China) ap-hongkong	Hong Kong Zone 1 (Nodes in Hong Kong, China can cover Hong Kong/Macao/Taiwan regions) ap-hongkong-1
	Hong Kong Zone 2 (Nodes in Hong Kong, China can cover Hong Kong/Macao/Taiwan regions) ap-hongkong-2

Other countries and regions

Region	AZ
Southeast Asia (Singapore) ap-singapore	Singapore Zone 1 (Singapore nodes can cover Southeast Asia Pacific) ap-singapore-1

	Singapore Zone 2 (Singapore nodes can cover Southeast Asia Pacific) ap-singapore-2
Northeast Asia (Seoul) ap-seoul	Seoul Zone 1 (Seoul nodes can cover Northeast Asia Pacific) ap-seoul-1
Northeast Asia (Tokyo) ap-tokyo	Tokyo Zone 1 (Tokyo nodes can cover Northeast Asia Pacific) ap-tokyo-1
South Asia Pacific (Mumbai) ap-mumbai	Mumbai Zone 1 (Mumbai nodes can cover South Asia Pacific) ap-mumbai-1
	Mumbai Zone 2 (Mumbai nodes can cover South Asia Pacific) ap-mumbai-2
Southeast Asia (Bangkok) ap-bangkok	Bangkok Zone 1 (Bangkok nodes can cover Southeast Asia Pacific) ap-bangkok-1
North America (Toronto) na-toronto	Toronto Zone 1 (Tokyo nodes can cover North America) na-toronto-1
Western US (Silicon Valley) na-siliconvalley	Silicon Valley Zone 1 (Silicon Valley nodes can cover Western US) na-siliconvalley-1
	Silicon Valley Zone 2 (Silicon Valley nodes can cover Western US) na-siliconvalley-2
Eastern US (Virginia) na-ashburn	Virginia Zone 1 (Virginia nodes can cover Eastern US) na-ashburn-1
	Virginia Zone 2 (Virginia nodes can cover Eastern US) na-ashburn-2
Europe (Frankfurt) eu-frankfurt	Frankfurt Zone 1 (Frankfurt nodes can cover Europe) eu-frankfurt-1
Europe (Moscow) eu-moscow	Moscow Zone 1 (Moscow nodes can cover Europe) eu-moscow-1

How to select the region and availability zone

When selecting a region and availability zone, take the following into consideration:

- Your location, the location of your users, and the region of the CVM instances.
We recommend that you choose the region closest to your users when making the CVM purchase

decision to minimize access latency and improve access speed.

- Other Tencent Cloud services you use.

When you select other Tencent Cloud services, we recommend you try to locate them all in the same region and availability zone, so they can communicate with each other through the private network to reduce access latency and increase access speed.

- High availability and disaster recovery.

Even if you have just one VPC, we still recommend that you deploy your businesses in different availability zones to prevent a single point of failure and enable cross-AZ disaster recovery.

- There may be higher network latency among different availability zones. We recommend that you assess your business requirements and find the optimal balance between high availability and low latency.
- If you need access to CVM instances in other countries or regions, we recommend you select a CVM in those other countries or regions. If you use a CVM instance in [Mainland China](#) to access [servers in other countries and regions](#), you may encounter much higher network latency.

Resource Availability

The following table describes which Tencent Cloud resources are global, which are regional, and which are specific to availability zones.

Resource	Resource ID Format 8-Digit String of Numbers and Characters	Type	Description
User Account	No limit	Globally unique	Users can use the same account to access Tencent Cloud resources from around the world.
SSH Keys	skey-xxxxxxx	Global	Users can use an SSH key to bind a CVM in any region under the account.
CVM Instances	ins-xxxxxxx	Specific to an availability zone	A CVM instance created in an availability zone is not available to other availability zones.
Custom Images	img-xxxxxxx	Available in multiple availability	Custom images created for the instance are available to all availability zones of the same region. Use **Copy Image** to copy a custom image if you need to use it in other regions.

		zones of a single region	
Elastic IPs	eip-xxxxxxx	Available in multiple availability zones of a single region	EIPs can only be associated with instances in the same region.
Security Groups	sg-xxxxxxx	Available in multiple availability zones of a single region	Security Group can only be associated with instances in the same region. Tencent Cloud automatically creates three default security groups for users.
Cloud Block Storage	disk-xxxxxxx	Specific to an availability zone	Users can only create a Cloud Block Storage disk in a specific availability zone and can mount it to instances in the same availability zone.
Snapshots	snap-xxxxxxx	Available in multiple availability zones of a single region	After creating a snapshot for a specific cloud disk, users can use this snapshot in this region for other operations (such as creating cloud disks).
Cloud Load Balancer	clb-xxxxxxx	Available in multiple availability zones of a single region	Cloud Load Balancer can be bound with CVMs in different availability zones of a single region for traffic forwarding.
VPC	vpc-xxxxxxx	Available in multiple availability zones of a single region	A VPC is created under a certain region. You can create resources in different AZs but in the same VPC.
Subnets	subnet-xxxxxxx	Specific to an availability zone	Users cannot create subnets across availability zones.
Route Tables	rtb-xxxxxxx	Available in multiple availability zones of a single region	When creating a routing table, users need to specify a VPC. Therefore, routing tables are regional as well.

Relevant Operation

Migrating an instance to another availability zone

Once launched, an instance cannot be migrated. However, you can create a custom image of your CVM instance and use the image to launch or update an instance in a different availability zone.

1. Create a custom image from the current instance. For more information, see [Creating Custom Images](#).
2. If the instance is on a [VPC](#) and you want to retain its current private IP address after the migration, first delete the subnet in the current availability zone and then create a subnet in the new availability zone with the same IP address range. Note that a subnet can be deleted only when it contains no available instances. Therefore, all the instances in the current subnet should be migrated to the new subnet.
3. Create a new instance in the new availability zone by using the custom image you have just created. You can choose the same type and configuration as the original instance, or choose new settings. For more information, see [Creating Instances via CVM Purchase Page](#).
4. If an elastic IP is associated with the original instance, dissociate it from the old instance and associate it with the new instance. For more information, see [EIPs](#).
5. (Optional) If the original instance is [pay-as-you-go](#), you can choose to terminate it. For more information, see [Terminating Instances](#). If the original instance is [monthly subscription](#), you can wait until it expires and then let the system repossess it.

Copying images to other regions

Operations such as launching and viewing instances are region-specific. If the image of the instance that you need to launch does not exist in the region, the image needs to be copied to the desired region. For more information, see [Copying Images](#).

Quota Limits for Cluster Purchase

Last updated : 2020-04-01 17:42:10

Each user is assigned a fixed quota for TKE clusters in each region.

TKE quota limits

The following describes the number of container clusters each user can purchase. If you need more clusters, [submit a ticket](#).

Since October 21, 2019, the maximum number of nodes supported in a cluster has been increased to 5,000.

Quota Item	Default Value	Where to Check	Support Quota Increase
Clusters in a single region	5	Bottom-right section of the TKE overview page	Yes
Nodes in a single cluster	5,000		
Image namespaces in a single region	10		
Images in a single region	500		
Image tags for a single image	100		

CVM quota limits

For CVM instances that you purchase for Tencent Cloud TKE, CVM purchase limits apply. For more information, see [CVM Instance Purchase Limits](#). See the following table for the maximum number of CVMs that a user can purchase by default. If you need a higher quota for any item, [submit a ticket](#).

Quota Item	Default Value	Support Quota Increase
Pay-as-you-go CVMs in a single availability zone	30 or 60	Yes

Cluster configuration limits

Cluster configuration limits the size of clusters and cannot be modified currently.

Configuration Item	IP Address Range	Affected Item	Where to Check	Support Modification
VPC network - Subnet	Custom	Number of nodes that can be added	VPC subnet list page for the cluster - Available IP addresses	<ul style="list-style-type: none">• No• Yes, and you can create new subnets
CIDR block of the container IP range	Custom	<ul style="list-style-type: none">• Maximum nodes per cluster• Maximum services per cluster• Maximum pods per node	Basic information page for the cluster - Container IP address range	No

Network Settings for Containers and Nodes

Last updated : 2020-09-18 10:20:02

Configuring Cluster and Node Networks

Cluster network and container network are the basic attributes of a cluster. You can plan the network of a cluster by configuring the cluster network and container network.

- **Cluster network:** assign IP addresses within the node network address range to CVMs in the cluster. You can select a subnet of a VPC instance as the node network of the cluster. For more information, see [VPC Overview](#).
- **Container network:** assign IP addresses within the container network IP range to containers in the cluster. You can customize three major private IP ranges as the container network to automatically assign the CIDR range of an appropriate size to the Kubernetes services based on the maximum in-cluster service quantity you select, or automatically assign an IP address range of an appropriate size to each CVM in the cluster for assigning IP addresses to Pods based on the maximum Pod quantity per node you select.

Relationship between the cluster network and container network

- The IP ranges of the cluster network and container network cannot conflict.
- The IP ranges of container networks in different clusters in the same VPC cannot conflict.
- If the container network conflicts with the VPC route, the traffic will be preferentially forwarded within the container network.

Communication between the cluster network and other Tencent Cloud resources

- Containers in the same cluster can communicate with each other.
- Containers and nodes in the same cluster can communicate with one another.
- Containers in a cluster can communicate via the private network with resources in the same VPC, such as TencentDB, [TencentDB for Redis](#), and TencentDB for Memcached.

Container network description

1. Container CIDR: it indicates the IP range where in-cluster resources such as services and pods are located.

2. Maximum service quantity per cluster: this determines the size of CIDR assigned to the service.

Note :

The services (kubernetes, hpa-metrics-service, and kube-dns) are created in a TKE cluster by default, and there are also 2 sets of broadcast addresses and network numbers; therefore, the maximum number of services you can use per cluster is (serviceMax - 5).

3. Maximum pod quantity per node: it determines the size of CIDR assigned to each node.

Note :

The pods are created in a TKE cluster by default: kube-dns-xxxx, kube-dns-xxxx, and l7-lb-controller-xxxx.

The following three addresses cannot be assigned to pods on a node: network number, broadcast address, and gateway address. Therefore, the maximum pod quantity per node is (podMax - 3).

Calculation of nodes in a cluster

1. Node: worker nodes in a cluster
2. The quantity of nodes is calculated as follows:

(CIDR IP quantity - Maximum service quantity per cluster)/Maximum pod quantity per node

Note :

You can customize the CIDR, maximum service quantity per cluster, and maximum pod quantity per node as required on the cluster information configuration page when you create a cluster.

For example, you can **set** the CIDR **to** 172.16.0.0/16, maximum pod quantity per node **to** 64, and maximum **service** quantity per cluster **to** 256, as shown **in** the following figure.

Container Network ⓘ

CIDR . . . / [Instruction](#)

Max pods per node

Max services in the cluster

Up to **1020** nodes are allowed in the current container network configuration.

The address space of the CIDR is allocated as follows:

node1 cidr	node2 cidr	nodeN cidr	service ip
------------	------------	-------	------------	------------

- The last IP range of the CIDR is allocated to services, providing a total of 256 IP addresses.
- The remaining IP addresses of the CIDR is calculated as $2^{(32 - 16)} - 256 = 65280$.
- Each node requires 64 IP addresses. Therefore, the cluster supports a maximum of 1,020 nodes ($65280/64 = 1020$) under the current container network configurations.

Container Node Disk Settings

Last updated : 2019-09-02 16:35:15

Description

When creating or scaling a TKE cluster, you can set the type and size of the system disks and data disks of the container nodes to meet your actual business needs.

Suggestions

1. The directory of the container is stored in the system disk. We recommend creating a system disk with a capacity of 50 GB.
2. If you have specific requirements for the system disk, you can move the Docker's directory to the data disk when initializing the cluster.

Notes on the Public IP of a TKE Node

Last updated : 2020-01-02 16:06:31

If you don't want to avoid exposing your company's IP while accessing the public network, you can use Tencent Cloud [NAT Gateway](#). This document describes how to access the public network via an NAT gateway.

Public IP

When a cluster is created, public IPs are assigned to the nodes in the cluster by default. With these public IPs, you can:

- Log in to the nodes in the cluster.
- Access services on the public network.

Public Network Bandwidth

When a service is created on the public network, the public network CLB uses the bandwidth and traffic of the nodes. If the public network service is required, the nodes need to have public network bandwidth. You can choose not to purchase public network bandwidth if it is not needed.

NAT Gateway

The CVM instance is not bound to an EIP, and all the traffic accessing the internet is forwarded via an NAT gateway. In this way, the traffic accessing the internet of the instance is forwarded to the NAT gateway over the private network. This means that the traffic is not subject to the upper limit of public network bandwidth specified when you purchase the instance, and the traffic generated from the NAT gateway does not occupy the public network bandwidth egress of the instance. To access the internet via an NAT gateway, follow the steps below:

Step 1. Create an NAT gateway

1. Log in to the [VPC Console](#) and click [NAT Gateway](#) in the left sidebar.
2. On the NAT gateway management page, click **Create**.
3. In the **Create an NAT Gateway** window that pops up, enter the following parameters.
 - Gateway Name: Custom.

- Network: Select the VPC of the NAT gateway service;
 - Gateway Type: Select based on actual needs. The type of the gateway can be changed after it is created.
 - Outbound Bandwidth Cap: Set based on actual needs.
 - Elastic IP: Assign an EIP to the NAT gateway. You can choose an existing EIP or purchase a new one.
4. Click **Create** to complete the creation of the NAT gateway.

The rental fee of 1 hour will be frozen during the creation of the NAT gateway.

Step 2. Configure the route table associated with the subnet

After the NAT gateway is created, you need to configure the routing rules on the route table page in the VPC Console to redirect the subnet traffic to the NAT gateway.

1. Click **Route Table** in the left sidebar.
2. In the route table list, click the route table ID/name associated with the subnet that needs to access the internet.
3. In the "Routing Policy" section, click **+ New routing policies**.
4. In the **Add routing** page, enter the **Destination**, select **NAT gateway** for **Next Hop Type**, and select the ID of the created NAT gateway for **Next Hop**.
5. Click **OK**.

Now, the traffic generated when the CVM instance in the subnet associated with the route table accesses the internet will be directed to the NAT gateway.

Other Solutions

Solution 1. Use an EIP

The CVM instance is only bound with an EIP but does not use an NAT gateway. With this solution, all the traffic of the instance accessing the internet goes out through the EIP and is subject to the upper limit of public network bandwidth specified when you purchase the instance. The fees for accessing the internet are charged based on the billing method of the instance's network.

For more information, see [Elastic Public IP](#).

Solution 2. Use both an NAT gateway and an EIP

If both an NAT gateway and an EIP are used, all the traffic of the CVM instance accessing the internet is forwarded to the NAT gateway over the private network, and the response packets are returned to the instance through the NAT gateway. This means that the traffic is not subject to the upper limit of public network bandwidth specified when you purchase the instance, and the traffic generated by the NAT gateway does not occupy the public network bandwidth egress of the instance. If the traffic from the internet proactively accesses the EIP of the instance, the response packets of the instance are all returned through the EIP. In this case, the resulting outbound public network traffic is subject to the upper limit of public network bandwidth specified when you purchase the instance. The fees for accessing the public network are charged based on the billing method of the instance's network.

If the bandwidth package (BWP) feature is activated in your account, fees of the outbound traffic generated by the NAT gateway will be deducted from the BWP (which means the network traffic will not be repeatedly billed at 0.12 USD/GB). It is recommended that you limit the outbound bandwidth of the NAT gateway so as to avoid high BWP fees due to excessive outbound bandwidth.

TKE Security Group Settings

Last updated : 2020-09-10 10:22:00

Security is always of utmost importance. Tencent Cloud prioritizes security in product design and strictly requires that all products are fully isolated. The Tencent Cloud classic network provides multiple layers of security protection. Tencent Kubernetes Engine (TKE) also pays special attention to security. TKE selects [VPC](#) with richer network features as the underlying network. This document describes the best practice of using security groups in TKE, helping you select the most appropriate security group policy.

Security Groups

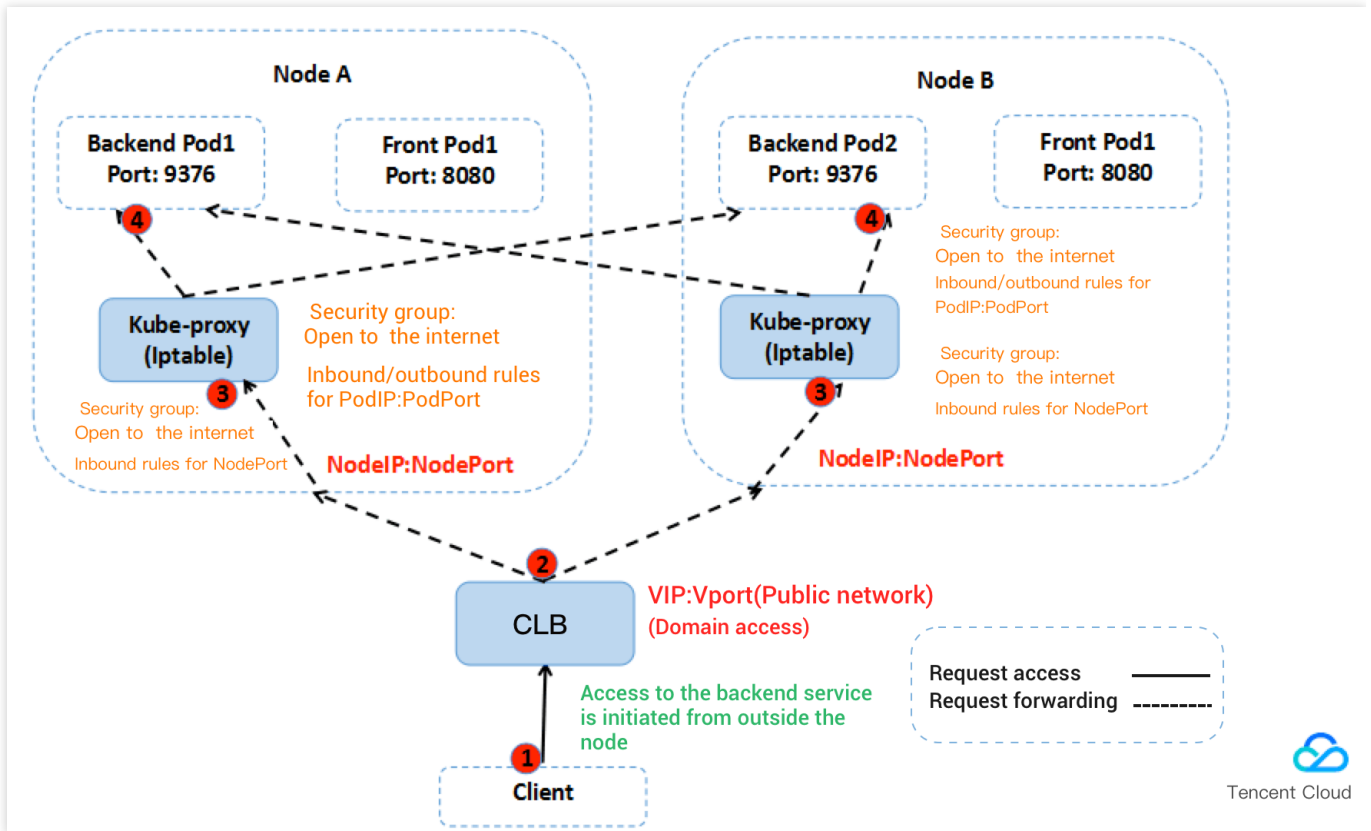
A security group is a virtual firewall for stateful data packet filtering. As an important network isolation approach provided by Tencent Cloud, a security group is used to set network access control of one or more Cloud Virtual Machine (CVMs). For more information on security groups, please see [Security Groups](#).

How to Select a Security Group for TKE

- In a container cluster, service pods are distributed on different nodes. We recommend that you bind all CVM instances in one cluster to the same security group and do not add non-clustered CVMs to a security group for a cluster.
- A security group only grants the minimum permission externally.
- You must enable the following rules for using TKE:
 - Open the container pod network and the cluster node network to the Internet.
When a node receives a service access request, the node forwards the request to a service pod according to the iptables rule configured by the kube-proxy module. If the service pod is on another node, cross-node access occurs. For example, the destination IP addresses of the access request include the IP address of the service pod, IP addresses of other nodes in the cluster, and the IP address of the cluster's cbr0 bridge on the node. In this case, the container pod network and the cluster node network on the peer node must be open to the Internet.
 - If clusters in the same VPC need to communicate with each other, you must open the container networks and node networks of the corresponding clusters to the Internet.
 - Open port 22 to the Internet if SSH login is required.
 - Open ports 30000 to 32768 on nodes to the Internet.
In the access path, you must use a load balancer to forward data packets to NodeIP:NodePort of

the container cluster. NodeIP is the CVM instance IP of any node in the cluster. NodePort is assigned by the container cluster by default when the service is created. NodePort ranges from 30000 to 32768.

The following figure uses service access from the public network as an example.



Default Security Group Rules for TKE

Some ports must be opened to the Internet to ensure normal communication between cluster nodes. To avoid cluster creation failures due to binding to invalid security groups, TKE provides default security group rules, as described in the following table.

⚠ Note :

If the current default security group cannot meet your service requirements and you have created a cluster bound to this security group, you can view and modify the security group rules for the cluster. For more information, please see [Managing Security Group Rules](#).

Inbound rules

Protocol	Port	Source IP	Rule	Description
----------	------	-----------	------	-------------

	Number	Address		
All	All	CIDR of the container network	Allow	Enable the communication between pods in the container network.
All	All	CIDR of the cluster network	Allow	Enable the communication between nodes in the cluster network.
TCP	22	0.0.0.0/0	Allow	Open the SSH login port to the Internet.
TCP	30000 to 32768	0.0.0.0/0	Allow	Enable the communication between master and worker nodes.
UDP	30000 to 32768	0.0.0.0/0	Allow	Enable the communication between master and worker nodes.
ICMP	-	0.0.0.0/0	Allow	Enable the support for Internet Control Message Protocol (ICMP) and ping operations.

Outbound rules

Protocol	Port Number	Source IP Address	Rule
All	All	0.0.0.0/0	Allow

Note :

- To customize outbound rules, you need to open the node IP range and container IP range.
- If you configure this rule for container nodes, the services in the cluster can be accessed using different access methods.
- For more information on how to access a service in a cluster, please see "Service Access" in [Overview](#).

Project of New Resources

Last updated : 2019-09-18 18:08:12

Overview

To conduct financial accounting by projects, please take the following into the consideration:

1. Clusters are not project-specific, but CVMs, load balancers and other resources in a cluster are project-specific.
2. Project of New-added Resource: Only resources newly added to the cluster are allocated to the project.

Notes

1. We recommend you allocate all the resources in a cluster to the same project.
2. If you need to distribute the CVMs in a cluster to different projects, go to the CVM Console to migrate projects.
3. If CVMs belong to different projects, they belong to different `security group instances` . Try to configure the same `security group rules` for the CVMs in the same cluster.

Purchase Elastic Clusters

Billing Overview

Last updated : 2020-04-28 18:47:37

Billing Mode

Elastic Kubernetes Service (EKS) is a pay-as-you-go service. The fees are calculated based on the configured amount of resources and the actual period of using them.

Billing Method

EKS calculates fees based on the specifications of the CPU, GPU, and memory for a workload and the running time of the workload. For more information, see [Product Pricing](#).

Other Fees

If you use EKS with other paid products such as [CLB](#), [CBS](#), and [CFS](#), these products are billed according to their own billing rules. For more information, see the purchase guide for the specific product.

Product Pricing

Last updated : 2020-09-11 10:10:39

Elastic Kubernetes Service (EKS) bills a pod by multiplying the [resource specification](#) of the pod by the **Unit price of the resource** and **Running time**. The following table shows the unit prices of the CPU and memory resources.

Billing Item	Original Price	Current Discount	Discount Price
CPU	0.000004976 USD per core per second	20%	0.000003981 USD per core per second
Memory	0.000002073 USD per GiB per second	20%	0.000001658 USD per GiB per second

The above offer is valid until 00:00:00 on May 15, 2020.

Star Lake AMD

Based on Tencent Cloud's self-developed Star Lake servers, EKS provides reliable, secure, and stable high performance. For more information, see [CVM Standard SA2 Introduction](#).

Billing Item	Price per Second	Price per Hour
CPU	0.00000219 USD per core per second	0.0079 USD per core per second
Memory	0.00000127 USD per GiB per second	0.0046 USD per GiB per second

Running Time

The running time is the time that elapses from when a pod fetches the first container image until the pod stops running. The pod is billed for the resources used during this period, which is measured in seconds.

Billing Examples

Sample 1

Assume that the specification of pods managed by a Deployment is 2 cores and 4 GB memory, and the number of replicas is fixed at 2. If the period from the time when the Deployment is launched to the time when the Deployment is terminated is 5 minutes, the Deployment is billed for the resources used during the running time of 300 seconds (5 minutes × 60 seconds).

In this case, the running fees of the Deployment = $2 \times (2 \times 0.000004976 + 4 \times 0.000002073) \times 300 = 0.019495$ USD.

Sample 2

Assume that a CronJob needs to launch 10 pods with 4 cores and 8 GB memory each time and terminate the pods 10 minutes later. If the CronJob executes the job twice a day and this task is managed by EKS, the task is billed as follows:

Daily task fees = $2 \times 10 \times (4 \times 0.000004976 + 8 \times 0.000002073) \times 600 = 0.437856$ USD.

Purchase Limits

Last updated : 2020-04-28 18:48:16

Prerequisites

Before using Elastic Kubernetes Service (EKS), you must complete identity verification for your Tencent Cloud account. For more information on identity verification, see [Identity Verification Guide](#).

Supported Regions

For more information on the regions and resource specifications supported by EKS, see [Resource Specifications](#).

Resource Quotas

Resource	Limit	Description
Clusters in one region	5	Includes clusters that are being created and running.
Pods in one cluster	50	Includes all namespaces, workloads, and stateless and stateful pods.
Pod replicas for one workload	50	Includes all stateless and stateful pods in the workload.

If the number of required resources exceeds the quota limit shown in the preceding table, you can submit a ticket to apply for a higher quota. Tencent Cloud will assess your actual needs and increase your quota as appropriate.

Applying for a higher quota

1. Log in to the [ticket system console](#). On the **Submit ticket** page, select **Other Problems** and then click **Create Now** to go to the page for creating a ticket.
2. In the **Problem description** field, enter a description such as "I want to apply for a higher quota for the elastic cluster." Then, enter the region where the elastic cluster is located and your desired quota. Finally, enter your mobile number and other information as instructed.
3. After providing all the necessary information, click **Submit ticket** to submit the ticket.

Resource specification

Last updated : 2020-11-16 14:34:20

Overview

Elastic Kubernetes Service (EKS) frees you from managing cluster nodes. However, to properly allocate resources and accurately calculate fees, you need to specify resource specifications for pods when deploying a workload. Tencent Cloud allocates computing resources to the workload and calculates the corresponding fees based on the specified specifications.

When you use the Kubernetes API or Kubectl to create a workload for EKS, you can use annotations to specify resource specifications. If annotations are not used, EKS will calculate the specifications based on the container parameters set for the workload, such as `Request` and `Limit`. For more information, see [Specifying Resource Specifications](#).

Note :

- Resource specifications indicate the maximum amount of resources available for containers in a pod.
- The following tables list the supported CPU and GPU specifications. Ensure that allocated resources do not exceed the supported specifications.
- The total amount of resources specified by `Request` for all the containers in a pod cannot exceed the maximum pod specification.
- The amount of resources specified by `Limit` for any container in a pod cannot exceed the maximum pod specification.

CPU Specifications

The following table lists the CPU specifications that EKS provides for pods in all regions where CPU resources are supported. EKS also provides a set of CPU options. Different CPU sizes correspond to different memory ranges. Select the CPU specification as needed when creating a workload.

Intel

CPU (Cores)	Memory Range (GiB)	Granularity of Memory Range (GiB)
0.25	0.5, 1, 2	-

0.5	1, 2, 3, 4	-
1	1 - 8	1
2	4 - 16	1
4	8 - 32	1
8	16 - 32	1
12	24 - 48	1
16	32 - 64	1

Star Lake AMD

The AMD processor provides high performance with high reliability, security, and stability based on the Star Lake servers developed by Tencent Cloud. For more information, see [Standard SA2](#).

CPU (Cores)	Memory Range (GiB)	Granularity of Memory Range (GiB)
1	1 - 4	1
2	2 - 8	1
4	4 - 16	1
8	8 - 32	1
16	16 - 32	1
32	32 - 64	1

GPU Specifications

The following table lists the GPU specifications that EKS provides for pods. Different GPU card models and sizes map to different CPU and memory options. Select the GPU specification as needed when creating a workload.

Note :

If you create, manage, and use GPU workloads using a YAML file, see [Annotation](#).

GPU Model	GPU (Cards)	CPU (Cores)	Memory (GiB)
-----------	-------------	-------------	--------------

1/4 Tesla V100-NVLINK-32G	1	2	10
1/2 Tesla V100-NVLINK-32G	1	4	20
Tesla V100-NVLINK-32G	1	8	40
Tesla V100-NVLINK-32G	2	18	80
Tesla V100-NVLINK-32G	4	36	160
Tesla V100-NVLINK-32G	8	72	320
1/4 NVIDIA T4	1	4	20
1/2 NVIDIA T4	1	10	40
NVIDIA T4	1	20	80
NVIDIA T4	2	40	160
NVIDIA T4	4	80	320

Supported Regions

The following table describes the regions, availability zones, and resource types that are currently supported by EKS.

Region	Availability Zone	Supported Resource Type
South China (Guangzhou) ap-guangzhou	Guangzhou Zone 3 ap-guangzhou-3	Intel CPU, GPU/vGPU (V100)
	Guangzhou Zone 4 ap-guangzhou-4	AMD CPU, GPU/vGPU (V100)
East China (Shanghai) ap-shanghai	Shanghai Zone 2 ap-shanghai-2	Intel CPU, GPU/vGPU(T4)
	Shanghai Zone 3 ap-shanghai-3	Intel CPU
	Shanghai Zone 4 ap-shanghai-4	Intel CPU
	Shanghai Zone 5 ap-shanghai-5	Intel/AMD CPU

East China (Nanjing) ap-nanjing	Nanjing Zone 1 ap-nanjing-1	Intel/AMD CPU
North China (Beijing) ap-beijing	Beijing Zone 3 ap-beijing-3	Intel/AMD CPU
	Beijing Zone 4 ap-beijing-4	Intel CPU

Billing Description

Last updated : 2020-11-13 15:53:21

TKE Billing Description

The TKE service itself is currently free of charge, but fees are collected for the use of related Tencent Cloud products. Use of TKE may require the use of the following products. For more information, see the billing documents of corresponding products.

- [CVM pricing modes](#)
- [CBS price overview](#).
- [Cloud Load Balancer Billing Description](#)

Note :

TKE is a Kubernetes-based declarative service. If you have created IaaS resources such as CLB and CBS in TKE and no longer need to use them, please delete the corresponding service resources in the TKE Console. If you only delete the service resources in the specific service Console, TKE will recreate them and fees will continue to be incurred. For example, if you have created CLB resources in TKE, and have deleted the CLB instance in the CLB Console, TKE will recreate the CLB instance based on the declarative API.

EKS Billing Description

EKS is pay-as-you-go and billed based on the configured resources and usage duration. For more information, see [Elastic Cluster Billing Overview](#).

Purchase Channels

Last updated : 2020-04-27 10:13:46

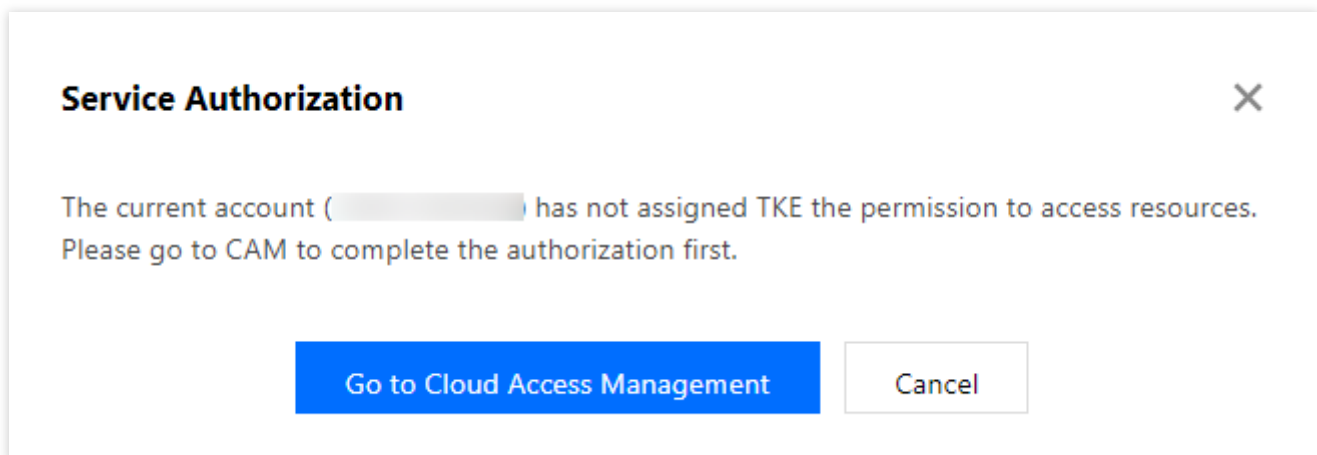
Product Purchase

Log in to the [purchase page of Tencent Cloud TKE](#). When logging in to this page for the first time, you need to perform the following steps to complete service authorization before you can purchase TKE products.

Service Authorization

Ignore the following procedure if you have completed service authorization.

1. View information in the displayed "Service Authorization" dialog box, and click **Go to Cloud Access Management**.



2. On the "Role Management" page, read information related to the role.

← **Role Management**

Service Authorization

After agreeing to grant permission for **TencentCloud Kubernetes Engine**, a preset role will be created for the service and relevant permissions will be granted to **TencentCloud Kubernetes Engine**

Role Name: TKE_QCSRole

Role Type: Service Role

Description: Current role is **TencentCloud Kubernetes Engine** Service role, which will access your other cloud service resources within the permissions of the associated policy

Authorized Policy: Preset policy QcloudAccessForTKERole ⓘ.

3. Click **Agree to Authorize** to grant authorization. Now, you can go to the [purchase page of Tencent Cloud TKE](#) to purchase related products.