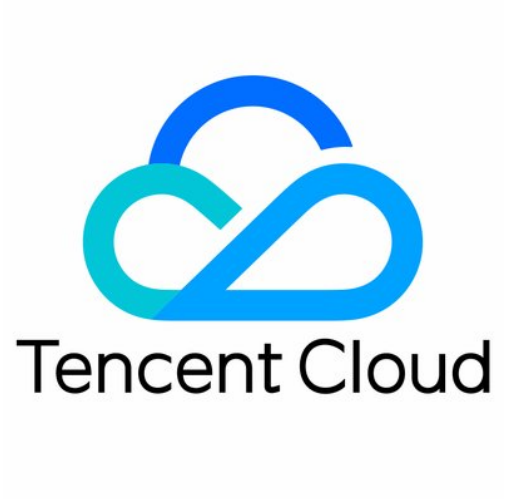
Cloud Application Rendering

Purchase Guide

Product Documentation



Contents

Purchase Guide
Billing Overview

Last updated：2024-03-20 09:06:44

CAR is billed by cloud application rendering concurrency. In CAR, a concurrency represents the collection of virtual computing resources, including CPU, bandwidth, disk, and GPU, required for one user to render your application content. Each concurrency supports access to your application by only one user at a time.

Billing items - CAR concurrency

A concurrency represents the collection of virtual computing resources, including CPU, bandwidth, disk, and GPU, required for one user to render your application content. Each concurrency supports access to your application by only one user at a time.
CAR supports hybrid scheduling of concurrencies in x86 and ARM architectures.
x86 concurrency: Has a high performance with the computing power of a desktop graphics card and is generally suitable for PC desktop applications.
ARM concurrency (in beta test): Has a high cost performance and low application adaptation costs and is generally suitable for Android applications.

Concurrency Architecture	Concurrency Scale	Application
x86	S - For rendering small applications	Small desktop applications
	M - For rendering medium-sized applications	Medium-sized desktop applications
	L - For rendering large applications	Large desktop applications
ARM	ARM concurrency (this feature is in beta test; if you want to use it, please contact your sales rep)	Android applications

Note：
In CAR, **a concurrency** is a virtual resource that supports one user to access your cloud application. **Concurrency scales** differ in their architecture and configuration, but they do not determine the number of concurrent users that can access an application. For example:
Suppose you have a large-scale virtual concert application. To guarantee there are enough computing resources to render your application, you can choose concurrency scale L with a high GPU performance. If you want to sustain 1,000 users accessing your cloud environment at the same time during peak hours, you can purchase 1,000 monthly

or daily subscribed L concurrencies and make additional users queue up to wait for concurrencies to become available. To learn more about queueing, see Queue Feature.
For more information on CAR concurrency scales, see Billing.

Billing mode

CAR concurrency packs are **prepaid** in a **monthly or daily subscribed manner or purchased as resource packages**. For more information, see Billing.
Note：
If you terminate and return a monthly subscribed CAR concurrency pack, it will be billed at the price for daily subscription.
A prepaid resource package is billed hourly. It takes effect upon purchase and is valid for six months, after which unused resources will expire and cannot be refunded or renewed. You can return an unused resource package for a full refund.
For more information, see Refunds.

Value-Added Feature Billing Item: Cloud-Based Push Streaming

CAR supports additional push streaming of the cloud-rendered images. There are two methods of push streaming: (1) Streaming to cloud live broadcasting by binding a cloud live broadcast domain; (2) Transmitting the target streaming address to stream the image to the specified address. The cloud-based push streaming feature needs to be enabled in the console before it can be called through the API. For detailed billing information, please refer to the Billing.

Value-Added Feature Billing: Multiplayer Interaction

In the multiplayer interaction mode, the room is created by a CAR player (i.e., the room owner), and then other players (i.e., interactive audience) can join the same room via the room owner's UserId. In the same room, all users can see the same cloud-rendered scene through a cloud rendering connection. To use the multiplayer interaction feature, it must be actively enabled in the console, and then it can be called via API. For detailed billing information regarding this value-added feature, please refer to billing.

Billing

Last updated：2024-03-20 09:08:05

CAR enables you to run your application in the cloud, so that end users can use your application in real time through a webpage or lightweight Android and iOS application. CAR is billed based on the concurrencies you purchase. One concurrency represents the collection of virtual computing resources, including CPU, bandwidth, disk, and GPU, required for one user to render your cloud application.

CAR Concurrency Scales

CAR supports hybrid scheduling of concurrencies in x86 and ARM architectures.
x86 concurrency: It has a high performance and computing power of a desktop graphics card, and is generally suitable for PC desktop applications.
ARM concurrency (in beta test): Has a high-cost performance and low application adaptation costs and is generally suitable for Android applications.

Concurrency Architecture	Concurrency Scale	Configuration					Applicatior
		CPU	Memory	GPU	Video Memory	Bandwidth	
x86	S - For rendering small applications	4-core or above vCPU performance	8 GB or above	2 TF SP/30T INT or above	4 GB or above	Up to 6 Mbps	Small desktop application
	M - For rendering medium-sized applications	4-core or above vCPU performance	16 GB or above	4 TF SP/30T INT or above	6 GB or above	Up to 8 Mbps	Medium-sized desktop application
	L - For rendering large applications	10-core or above vCPU performance	32 GB or above	8.1 TF SP/30T INT or above	12 GB or above	Up to 8 Mbps	Large desktop application
ARM (in beta test)	ARM concurrency (this feature is in beta test and available in multiple scales. If you want to use it, please contact your Tencent Cloud sales rep)						Android application

Note:
The ARM concurrency feature is in beta test. If you want to use it, please contact your Tencent Cloud sales rep.
In CAR, **a concurrency** is a virtual resource that supports one user to access your cloud application. **Concurrency scales** differ in their architecture and configuration, but they do not determine the number of concurrent users that can access an application.
Suppose you have a large-scale virtual concert application. To guarantee there are enough computing resources to render your application, you can choose concurrency scale L with a high GPU performance. If you want to sustain 1,000 users accessing your cloud environment at the same time during peak hours, you can purchase 1,000 monthly or daily subscribed L concurrencies and make additional users queue up to wait for concurrencies to become available. To learn more about queueing, see Queue Feature.
The main metrics for the GPU performance are floating-point operations capabilities.
TF indicates floating-point operations per second (FLOPS).
SP indicates single-precision floating-point operations.
DP indicates double-precision floating-point operations.
INT8 indicates INT8 integer operations.

Concurrency Billing Modes

Currently, CAR concurrency is available in three prepaid billing modes: monthly subscription, daily subscription, and prepaid resource packages.
Prepaid monthly subscription: Suitable for **a business with stable traffic** or **multiple businesses sharing resources**, such as virtual branches, shops, and exhibition halls, which have steady user access.
Prepaid daily subscription: Suitable for **virtual events** such as virtual concerts and cloud marketing events, which involve a high concurrency.
Prepaid resource package: Hourly billing and the hours are purchased before being used. This billing mode is suitable for businesses that have **irregular concurrent user access and require elastic scaling**.
We recommend that you select an optimal billing mode based on your business scenario. If you have special requirements for billing, please contact your Tencent Cloud sales rep for assistance.

Prepaid monthly and daily subscription

Monthly subscription

Monthly subscriptions are suitable for a business with stable traffic or multiple businesses sharing resources, such as virtual branches, shops, and exhibition halls, which have steady user access.

Concurrency Architecture	Concurrency Scale	AZ					Billing Cycle	Billin
		North America	Singapore	Tokyo	Seoul	Frankfurt		

x86 (high performance)	S - For rendering small applications	✓	✓	✓	✓	✓	Monthly	USD/ user/
	M - For rendering medium-sized applications	✓	✓	✓	✓	✓		
	L - For rendering large applications	✓	✓	✓	✓	✓		
ARM (high cost-effectiveness)	ARM concurrency (in beta test)			✓				

Daily subscription

Daily subscriptions are suitable for virtual events such as virtual concerts and cloud marketing events, which involve a high concurrency.

Concurrency Architecture	Concurrency Scale	AZ					Billing Cycle	Billing U
		North America	Singapore	Tokyo	Seoul	Frankfurt		
x86	S - For rendering small applications	✓	✓	✓	✓	✓	Daily	USD/co user/da
	M - For rendering medium-sized applications	✓	✓	✓	✓	✓		
	L - For rendering large applications	✓	✓	✓	✓	✓		

Note:

The ARM concurrency feature is in beta test and cannot be purchased in the console. If you want to use it, please contact your Tencent Cloud sales rep.

Fee calculation method for monthly or daily subscription

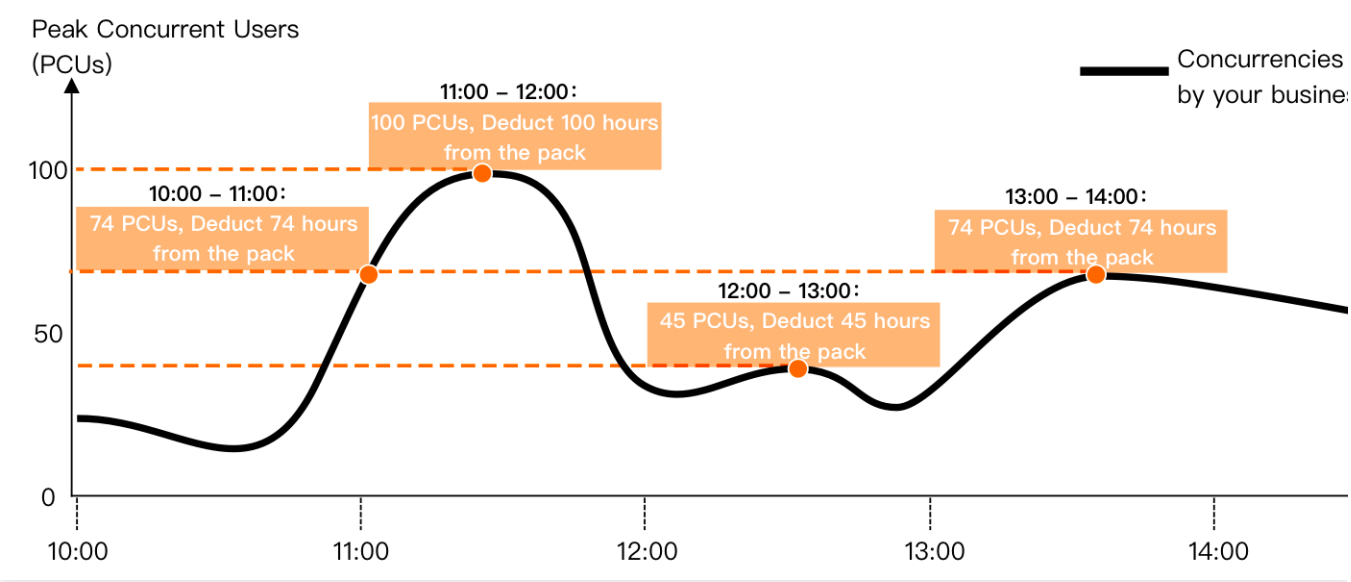
Fees = Unit price of the specified concurrency scale x Number of concurrencies x Purchase duration

Example:

Suppose you need to hold a cloud exhibition for one month and want to sustain up to 100 concurrent users accessing your exhibition hall application on the first day of the exhibition and up to 10 concurrent users on the later days. For this, you can purchase a daily subscribed concurrency pack with 90 concurrencies for one day and a monthly subscribed concurrency pack with 10 concurrencies for one month (for additional users, you can use the queue feature to utilize resources efficiently). If the unit prices of scale S in Singapore region are 100 USD/month and 10 USD/day, the total fees will be 10 x 90 x 1 + 100 x 10 x 1 = 1900 USD.

Prepaid resource package

A prepaid resource pack is a billing mode where CAR resources are purchased before being used. A resource pack can be used for applications that experience irregular concurrent access and require elastic scaling. A prepaid resource package is billed **hourly**. The system counts the **peak number of concurrencies** in an hour and deducts it from the total number of hours included in the package.



Example:

Suppose you purchased a prepaid resource package of 10,000 hours and there were up to 25, 10, and 74 concurrent users in the first, middle, and last 20 minutes respectively between 10:00 and 11:00, the peak number of

concurrencies in the hour is 74. After 11:00, CAR will settle the fees for the last hour and deduct 74 hours from the package of 10,000 hours, and there will be 9,926 hours available remaining.

Note：

Concurrencies under a prepaid package **do not support prelaunch**, and even if they are allocated to a single-application project as described in Project types and the prelaunch feature is enabled for the project, prelaunch will not take effect if a prepaid package is used. To quickly start the application without loading it, you need to purchase a monthly or daily subscribed concurrency pack.

It takes time to add concurrencies under a prepaid resource package, and up to 50 concurrencies can be added per minute. If too many users access your application faster than concurrencies are added, a failure will be prompted. Therefore, we recommend that you purchase daily subscribed concurrency packs for scenarios involving a spike in concurrencies.

Multiple prepaid resource packs can be bound to the same project or different projects. However, **they do not support accumulating concurrency limits. The actual available concurrency depends on real-time inventory.** When the monthly or daily demands are high and result in insufficient inventory, the resource packs may not provide the maximum concurrency, and users may need to queue.

Resource pack specifications and AZs

<table>
<tr><th colspan="5">Resource Pack Specification</th><th rowspan="2">AZ</th></tr>
<tr><th>Concurrency Architecture</th><th>Concurrency Scale</th><th>Total Hours</th><th colspan="2">Peak Concurrency Limit</th></tr>
<tr><td rowspan="8">x86</td><td rowspan="4">S - For rendering small applications</td><td>10,000 hours</td><td>≤500 concurrency</td><td rowspan="8">Regardless of whether they are bound to the same project or different projects, the peak concurrency limits of multiple resource packs cannot accumulate.

The actual available concurrency depends on real-time inventory. When the monthly or daily demands are high and</td><td rowspan="8">North America, Singapore, Tokyo, Seoul, and Frankfurt

(If resource packs from different regions are required, they must be purchased separately.)</td></tr>
<tr><td>5,000 hours</td><td>≤400 concurrency</td></tr>
<tr><td>2,000 hours</td><td>≤200 concurrency</td></tr>
<tr><td>1,000 hours</td><td>≤100 concurrency</td></tr>
<tr><td rowspan="4">M - For rendering medium-sized applications</td><td>10,000 hours</td><td>≤500 concurrency</td></tr>
<tr><td>5,000 hours</td><td>≤400 concurrency</td></tr>
<tr><td>2,000 hours</td><td>≤200 concurrency</td></tr>
<tr><td>1,000 hours</td><td>≤100</td></tr>
</table>

			concurrency	result in insufficient inventory, the resource packs may not provide the maximum concurrency. The peak concurrency of 500 is a limit set for the resource pack. If the demand for concurrency is more than 500, you can make multiple purchases of monthly/daily resource packs, or contact us to evaluate special billing models and quotes.	
	L - For rendering large applications	10,000 hours	≤500 concurrency		
		5,000 hours	≤400 concurrency		
		2,000 hours	≤200 concurrency		
		1,000 hours	≤100 concurrency		
	XL - For rendering extra large applications	10,000 hours	≤500 concurrency		
		5,000 hours	≤400 concurrency		
		2,000 hours	≤200 concurrency		
		1,000 hours	≤100 concurrency		

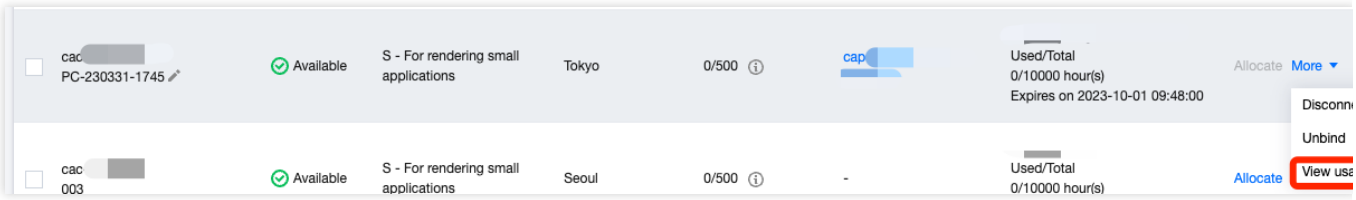
Note

Purchasing instructions for prepaid resource packs

After purchasing, the prepaid resource pack becomes immediately effective and remains valid for 6 months. Unused resource packs will expire at the end of this period. Validity period of each resource pack is calculated separately, and that of multiple ones do not accumulate.

Prepaid resource packs that satisfy the requirement for a five-day unconditional refund are eligible for refunds. However, under other circumstances, once activated, these packs will not be eligible for refunds.

To inquire about the activation and remaining amount of resource packs, please go to [Concurrency Management - View Usage] in the console.



How to understand the concurrency upper limit of the prepaid resource packs? What are the differences between prepaid resource packs and directly purchasing monthly or daily concurrency packs?

You can exclusively use the monthly and daily packs, while the concurrency in the resource packs is called in real-time from the inventory. Therefore, the concurrency upper limit in the resource packs is the theoretical maximum, the actual supply of concurrency depends on the real-time inventory. For instance, you have purchased a resource pack containing 10,000 hours, which supports up to 500 users connecting simultaneously. However, if the demand for monthly and daily packs is high one day resulting in only 100 remaining concurrency instances in the inventory, then the resource packs can only support 100 users at most at the same time, the others will have to wait in a queue.

Multiple prepaid resource packs do not support accumulating in concurrency limit whether they are bound to the same project or different projects.

The peak concurrency of 500 is a limit set for the resource pack. If the demand for concurrency is more than 500, you can make multiple purchases of monthly/daily resource packs, or contact us to evaluate special billing models and quotes.

When a project is simultaneously bound with monthly&daily concurrency packs and resource packs, the monthly&daily concurrency packs will be prioritized. Usage exceeding the limit of the monthly&daily concurrency packs will be charged through the resource pack billing.

Apart from the billing differences, are there any functional differences between prepaid resource packs and prepaid monthly&daily concurrency packs?

Prepaid resource packs do not support pre-launching, meaning that even when assigned to Project: Single-application with the pre-launch feature enabled, the resource packs can not start pre-launching. To ensure an instant loading in the application, it is necessary to purchase monthly or daily concurrency packs.

Expansion of prepaid resource pack concurrency will take time. The maximum amount of concurrency is 50 per minute. Any attempt at exceeding this limit will result in failure and users will need to join a queue. Therefore, for scenarios involving high concurrency in a short time, it is recommended to purchase daily concurrency packs, or contact us to evaluate special billing models and quotes.

Special billing mode

We recommend that you select an optimal billing mode based on your business scenario as described in Selecting the Optimal Billing Mode Based on the Business Scenario. If neither monthly subscription nor daily subscription can meet the requirements for your specific business scenario, please contact your Tencent Cloud sales rep for a special billing mode and quotation.

Optimal Billing Mode by Scenario

Typical business scenarios

Type	Characteristics		Example	Recommended Billing Mode
Business with stable traffic	A single business	The daily peak number of concurrent users is stable, and there are no obvious peak and off-peak hours.	Scenarios with a stable daily concurrency such as a virtual shop and 3D automobile viewer	Monthly subscription

	Multiple businesses	The businesses are only loosely related, and the total daily peak number of their concurrent users is stable.	Multiple cloud exhibition hall projects unrelated to each other	
Business with long-term stable traffic and short-term spikes	A single business	The business has obvious peak and off-peak hours (high concurrency during certain times and stable traffic for the rest of the time).	Cloud marketing events, where the concurrency is high during the event and gradually decreases to a stable level after the event	Monthly subscription (for long-term stable traffic) + daily subscription (for short-term high concurrency)
	Multiple businesses	The businesses are loosely related. Certain businesses involve a high concurrency sometimes but stable traffic most of the time.	Multiple digital architectural model projects. Each project involves a high concurrency during the launch, but the total number of concurrent users for all projects is stable for the rest of the time.	
Business with a traffic spike during an event	Businesses with a short-term traffic spike during an event, which involve a high concurrency during a short period of time		Virtual concert or festival celebration, which requires over 1,000 or even 10,000 concurrencies during the event	Daily subscription

Special business scenarios

If you are uncertain of your business scenario or your business scenario has special requirements, please contact your Tencent Cloud sales rep for a special billing mode and quotation.

Value-Added Features: CAR Cloud-Based Streaming

CAR supports the additional streaming of images rendered in the cloud. Cloud-based streaming supports two ways of streaming: (1) Streaming to Tencent Cloud Streaming Services (CSS) by binding a cloud live broadcast domain; (2)

Transmitting the target streaming address, stream the image to the specified address.

Streaming to Tencent Cloud Streaming Services (CSS)

For the pricing of cloud live streaming products, please refer to the description Billing of LVB. No additional streaming fees are charged during the application of cloud rendering products.

Streaming to a Specified Address

Convey the target streaming address, directing the video stream to the specified address. When utilizing this feature, the system will record the peak bandwidth of each stream, and calculate the total peak bandwidth (Mbps) generated in the billing period streaming service region. The region where cloud rendering is concurrently used is the streaming region. If you stream to multiple regions in a billing period, fees will be charged separately based on the peak bandwidth usage in each region.

Region	Price (USD/Mbps/Month)
Chinese mainland	12.67
Singapore	8.04
Frankfurt	7.1
Seoul	16.56
Silicon Valley	7.1
Virginia	7.1
Japan	13.01

Billing details

Billing mode: Monthly pay-as-you-go
Billing cycle: Monthly billing. Your bill for each month is generated between the 1st and 3rd day of the following month.
Billing rules: Streaming to specified address fees are charged in the pay-as-you-go mode based on your average daily peak bandwidth usage in each month.

Billing examples

Suppose you streaming to a specified address on five days in August 2023. The region was the Chinese mainland, and the peak bandwidth used on the five days were 10 Mbps, 80 Mbps, 70 Mbps, 75 Mbps, and 60 Mbps respectively.
There are 31 days in August. Your streaming to specified address fee for August would be as follows:
`(10 Mbps + 80 Mbps + 70 Mbps + 75 Mbps + 60 Mbps) /31 days x 12.67 (USD/Mbps/month) = 120.569 USD` .

Value-Added Feature Billing Description: Multiplayer Interaction

In the multiplayer interaction mode, the room is created by a CAR player (i.e., the room owner), and then other players (i.e., interactive audience) can join the same room via the room owner's UserId. In the same room, all users can see the same cloud-rendered scene through a cloud rendering connection. To use the multiplayer interaction feature, it must be actively enabled in the console, and then it can be called via API.
When this feature is used, the system will record the bandwidth of interactive users. The bandwidth of multiple users will be cumulatively calculated to determine the peak value, and the peak bandwidth (in Mbps) generated in the service's region during the billing cycle will be the basis for settlement. The region of cloud-rendering concurrency usage will be considered the region for multiplayer interaction. If multiplayer interaction services occur in multiple regions in the same billing cycle, the bandwidth peak values for the involved regions will be billed separately.

Region	Price (USD/Mbps/Month)
Mainland China	12.67
Singapore	8.04
Frankfurt	7.1
Seoul (South Korea)	16.56
Silicon Valley, USA	7.1
Virginia, USA	7.1
Tokyo, Japan	13.01

Billing Details

Billing method: Postpaid monthly billing.
Billing cycle: billed monthly, with multiplayer interaction billing statements for the previous month generated from the 1st to the 3rd of the following month.
Billing rule: Billing is based on the total bandwidth of all users involved in multiplayer interaction, with the default postpaid mode being the average of daily peak values within the billing cycle (average daily peaks per month).
Note:
The bandwidth used by the host is not counted; the bandwidth of multiple interacting users will be cumulatively calculated. For example, for a multiplayer interaction concurrency with 1 host and 2 interactive players, each with a bandwidth of 5Mbps. During calculation, the host's bandwidth is excluded, and the bandwidth of the 2 interactive players is merged for statistics, resulting in 10Mbps.

Billing Example

The user utilized the cloud rendering multiplayer interaction feature for a total of 5 days in August 2023, in the region of Mainland China, with daily peak bandwidths of 10Mbps, 80Mbps, 70Mbps, 75Mbps, and 60Mbps. Given that the maximum number of days in August 2023 is 31 days, the fee for the multiplayer interaction value-added feature in August is:

`(10Mbps + 80Mbps + 70Mbps + 75Mbps + 60Mbps)/31 days x 12.67 (USD/Mbps/month) = 120.569 USD` .

Purchase Process

Last updated：2024-01-26 11:54:09

Before reading this document, you need to understand the basic concepts of CAR.

Prerequisites

Before purchasing a CAR concurrency pack, you need to perform the following steps:

1. Sign up for a Tencent Cloud account.
2. Submit a CAR integration application.

Directions

1. After the application is approved, enter the CAR console.
2. Go to the **Concurrency management** page, click **Create**, and set the configuration items as follows:

Configuration Item	Description
Application type	Cloud 3D by default.
Project	Allocate the concurrency pack to one of your projects, or select **To be allocated** to allocate it later. **After the concurrency pack is allocated to a project, you can still transfer it to another project later**.
Billing mode	Select the billing mode for the concurrency pack. Currently, CAR supports two prepayment modes: monthly and daily.
Region	Select the region of the concurrency pack based on the region of your business to ensure the optimal cloud rendering and operation experience.
Concurrency scale	Select the concurrency scale based on the concurrency scale you configured for your application. For more information on the supported concurrency scales, see Concurrency Scales.
Quantity	Select the number of concurrencies (i.e., the max number of users that can concurrently access your application). Each concurrency supports access by only one user at a time. You can select the number of concurrencies you want based on the peak and average numbers of concurrent users than you expect to access your application.
Duration	Select the amount of time that the CAR concurrency pack will be used.
Auto-renew	When **Auto-renew** is enabled, your concurrency pack will be automatically renewed

(optional)	monthly upon expiration if your Tencent Cloud account balance is sufficient. **Note: Concurrency packs billed on a daily basis don't support automatic renewal**.
Pack name	Name the CAR concurrency pack.

3. After completing the above configuration, confirm that everything is correct and click **Buy Now**.
4. On the **Check the order** page, check the details and pay for the order.
5. After making the payment, click **Go to the console** to manage your CAR concurrency pack.

Overdue Payment

Last updated：2024-01-26 11:54:09

Note:

If you are a customer of a Tencent Cloud partner, the rules regarding resources when there are overdue payments are subject to the agreement between you and the partner.

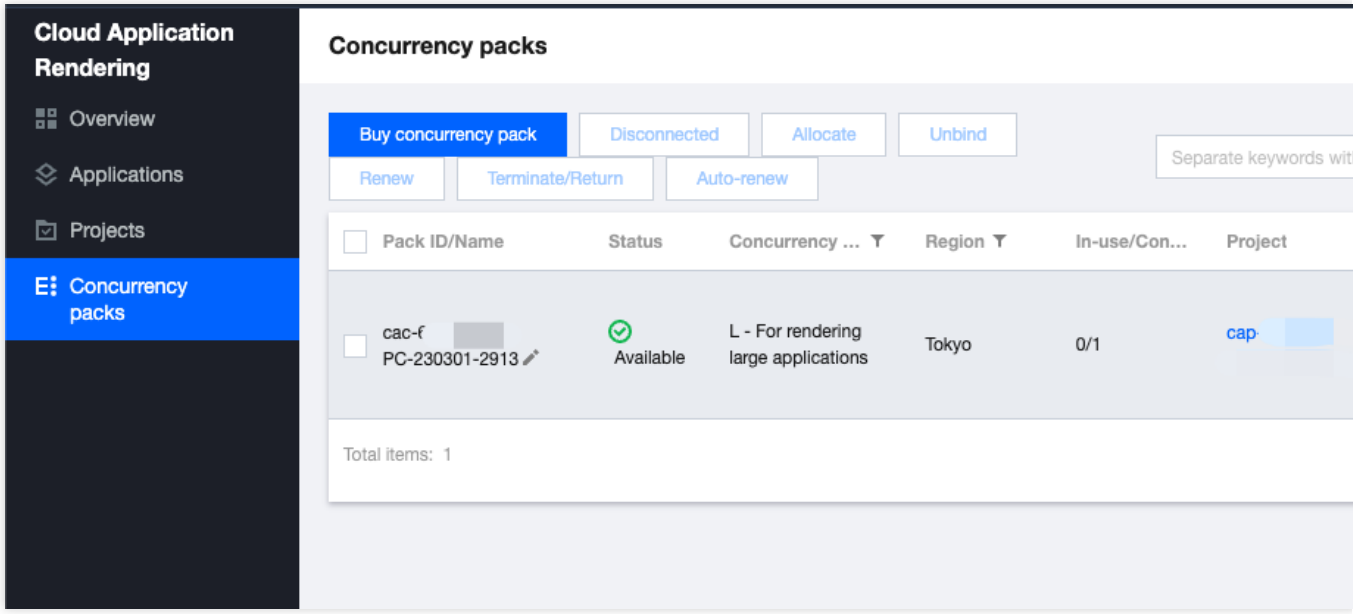
CAR supports the following two alert types:

Alert Type	Description
Expiration	Seven days before your CAR concurrency pack expires, Tencent Cloud will send an expiration alert to your Tencent Cloud account creator, global resource collaborators, and financial collaborators by email and SMS.
Overdue payment	On the day of and after CAR concurrency pack expiration, Tencent Cloud will send an overdue payment alert to your Tencent Cloud account creator and all collaborators by email and SMS.

Refunds

Last updated：2024-01-26 11:54:09

If you are not satisfied after purchasing a CAR concurrency pack, you can return it for a partial refund. Fees for consumed resources will be deducted from the refund, and **the remaining amount, including cash and free credit, will be credited to your Tencent Cloud account.** You can perform the above operations in the console.



Self-Service Return

Monthly and daily subscription

You can return up to 199 monthly subscribed CAR concurrencies in the console in a self-service manner. In standard return mode, the refund amount is calculated at the daily subscription price, that is, **Refund amount = Paid monthly subscription fees - Daily subscription price**.

Example

Suppose you purchased an L CAR concurrency pack in Tokyo region for one month in a monthly subscribed manner at 200 USD/month (price shown is for demonstration only) and you are dissatisfied and want to return it after 2 days and 10 hours of use, then:

Refund amount = actual payment amount - usage duration x daily subscription price = 200 - 20 x 3 = 140 USD (20 USD/day is the daily subscription price of an L concurrency pack. The price shown is for demonstration only)

Note:

The above prices are for demonstration purposes only and do not reflect the actual pricing shown on the official website. If you make a return, the actual unit prices at the time of purchase shall apply, which may vary by region,

promotional campaign, or policy.

Prepaid resource package

A prepaid resource package takes effect upon purchase and is valid for six months, after which it will expire without a refund and cannot be renewed.
You can return an unused resource package for a full refund.

Example

Suppose you purchased an S resource package of 10,000 hours for 20,000 USD (price shown is for demonstration only).
1. If you have not used the resource package and want to return it after several months, you can get a full refund of 20,000 USD for the remaining 10,000 hours.
2. If there are 9,000 hours remaining in the resource package after five days of use, the fees are non-refundable.
3. If there are 100 hours remaining in the package after six months of use, the fees are non-refundable and the package cannot be renewed, as the validity period of six months has expired.

Note

After you return a concurrency pack in a self-service manner, once its status becomes **Terminating**, it will no longer incur any fees.