

应用平台

# 计费说明

文档版本	03
发布日期	2024-04-16



版权所有 © 华为云计算技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

## 商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

## 注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

## 华为云计算技术有限公司

地址：贵州省贵安新区黔中大道交兴功路华为云数据中心 邮编：550029

网址：<https://www.huaweicloud.com/>

目 录

1 计费概述.....	1
2 计费模式.....	2
2.1 计费模式概述.....	2
2.2 包年/包月.....	3
2.3 按需计费.....	7
3 计费项.....	9
3.1 开发中心/运维中心/运行时引擎/运营中心计费项.....	11
3.2 AI 原生应用引擎计费项.....	14
4 续费.....	17
4.1 续费概述.....	17
4.2 手动续费.....	18
4.3 自动续费.....	20
5 费用账单.....	22
6 欠费说明.....	24
7 停止计费.....	26
8 计费 FAQ.....	28
9 修订记录.....	29

# 1 计费概述

通过阅读本文，您可以快速了解应用平台（AppStage）的计费模式、计费项、续费、欠费等主要计费信息。

- **计费模式**

- 应用平台的开发中心、运维中心、运行时引擎、运营中心提供包年/包月计费模式。包年/包月是一种预付费模式，即先付费再使用，按照订单的购买周期进行结算，因此在购买之前，您必须确保账户余额充足。
- 应用平台的AI原生应用引擎根据资源不同提供包年/包月和按需计费两种计费模式。

计费模式详细介绍请参见[计费模式](#)。

- **计费项**

应用平台开发中心、运维中心、运行时引擎、运营中心的计费项为套餐规格的包年/包月购买。关于计费项的详细介绍，请参见[计费项](#)。

- **续费**

包年/包月应用平台服务在到期后，会影响应用平台的正常运行。如果您想继续使用应用平台，需要在规定的时间内为应用平台进行续费。关于续费的详细介绍，请参见[续费](#)。

- **费用账单**

您可以在“费用中心 > 账单管理”中，查看与应用平台相关的流水和明细账单，以便了解您的消费情况。如需了解具体操作步骤，请参见[费用账单](#)。

- **欠费**

对于包年/包月的应用平台服务，用户已预先支付了版本费用，因此在账号出现欠费的情况下，已购买的应用平台版本仍可用。关于欠费的详细介绍，请参见[欠费说明](#)。

- **停止计费**

当应用平台不再使用时，可以退订或删除，从而避免继续收费。详细介绍请参见[停止计费](#)。

# 2 计费模式

## 2.1 计费模式概述

应用平台提供包年/包月和按需计费两种计费模式，以满足不同场景下的用户需求。

- 包年/包月：一种预付费模式，即先付费再使用，按照订单的购买周期进行结算。购买周期越长，享受的折扣越大。一般适用于设备需求量长期稳定的成熟业务。
- 按需计费：一种后付费模式，即先使用再付费，按云服务计算资源的实际使用时长计费，秒级计费，按小时结算。按需计费模式允许您根据实际业务需求灵活地调整资源使用，无需提前预置资源，从而降低预置过多或不足的风险。一般适用于资源需求波动的场景，可以即开即停。

表2-1列出了两种计费模式的区别。

表 2-1 计费模式

计费模式	包年/包月	按需计费
付费方式	预付费	后付费
计费周期	按订单的购买周期计费。	秒级计费，按小时结算。 按照实际使用时长计费。

适用计费项	<ul style="list-style-type: none"><li>● 适用开发中心/运维中心/运行时引擎/运营中心如下计费项：<ul style="list-style-type: none"><li>- 开发中心专业版（开发中心用户数）</li><li>- 开发人员增量包（开发中心用户数）</li><li>- 运维中心专业版（应用实例数）</li><li>- 运维中心实例增量包（应用实例数）</li><li>- 运维中心服务扩展包（日志数、span数、短信数）</li><li>- 运行时引擎专业版（应用实例数）</li><li>- 运行时引擎实例增量包（应用实例数）</li><li>- 运营中心专业版（事件量、看板数、指标数）</li><li>- 运营中心事件量增量包（事件量）</li><li>- 运营中心看板增量包（看板数）</li><li>- 运营中心指标增量包-10个指标</li><li>- 运营中心指标增量包-100个指标</li><li>- 运营中心指标管理增强包</li><li>- 运营中心数智化大屏增强包</li></ul></li><li>● 适用AI原生应用引擎如下计费项：<ul style="list-style-type: none"><li>- 问答AI服务基础版（大模型调用token数、用户数）</li><li>- 问答AI服务专业版（大模型调用token数、用户数）</li><li>- 推理单元包周期（推理单元数）</li></ul></li></ul>	<p>适用AI原生应用引擎如下计费项：</p> <ul style="list-style-type: none"><li>● 大模型微调服务API在线调用-SFT局部调优</li></ul>
适用场景	适用于可预估资源使用周期的场景，价格比按需计费模式更优惠。对于长期使用者，推荐该方式。	适用于云服务资源需求波动的场景，可以随时开通，随时删除。

## 2.2 包年/包月

包年/包月是一种先付费再使用的计费模式，适用于对资源需求稳定且希望降低成本的用户。通过选择包年/包月的计费模式，您可以预先购买应用平台并获得一定程度的价格优惠。本文将介绍包年/包月应用平台产品的计费规则。

适用场景

包年/包月计费模式需要用户预先支付一定时长的费用，适用于长期、稳定的业务需求。以下是一些适用于包年/包月计费模式的业务场景：

- 稳定业务需求：对于长期运行且资源需求相对稳定的业务，如企业官网、在线商城、博客等，包年/包月计费模式能提供较高的成本效益。
- 长期项目：对于周期较长的项目，如科研项目、大型活动策划等，包年/包月计费模式可以确保在整个项目周期内资源的稳定使用。
- 业务高峰预测：如果能预测到业务高峰期，如电商促销季、节假日等，可提前购买包年/包月资源以应对高峰期的需求，避免资源紧张。
- 数据安全要求高：对于对数据安全性要求较高的业务，包年/包月计费模式可确保资源的持续使用，降低因资源欠费而导致的数据安全风险。

适用计费项

应用平台以下计费项支持包年/包月。

表 2-2 计费项说明

套餐包类型	说明	计费项	规格约束
推理单元包 周期	将基础或者预训练后的模型部署为在线服务时使用的算力单元，通过购买更多推理单元实现服务QPS扩容。	推理单元数	<ul style="list-style-type: none"><li>● 6000/推理单元/月</li><li>● 60000/推理单元/年</li></ul>
开发中心专业版	<ul style="list-style-type: none"><li>● 统一应用架构建模</li><li>● 产品/版本/个人级协作管理</li><li>● 一站式开发工具链</li><li>● 研发效能洞察</li><li>● 海量资产模板</li></ul>	开发中心用户数	开发人员数量50人。超过50人可单独购买开发人员增量包。
开发人员增量包	开发中心用户数扩展	开发中心用户数	200元/用户/月

套餐包类型	说明	计费项	规格约束
运维中心专业版	<ul style="list-style-type: none"><li>● 视界监控服务</li><li>● 纪实日志服务</li><li>● 追踪者调用链服务</li><li>● 先知告警服务</li><li>● 智维AIOps</li><li>● 蓝军故障演练服务</li></ul> <p><b>说明</b> 运维中心不依赖于运行时引擎，可以只订购运维中心，不订购运行时引擎。</p>	应用实例数	<ul style="list-style-type: none"><li>● 应用实例数量20</li><li>● 100指标/实例</li><li>● 2G/ 天/实例</li><li>● 1万span/天/实例</li><li>● 5条语音或短信/天/实例</li></ul> <p>超过上述规格可单独购买实例增量包、服务扩展包。</p> <p><b>说明</b> 实例为应用部署节点，一个虚拟机或者容器化部署一个Pod为一个实例。</p>
运维中心实例增量包	运维中心应用实例数扩展	应用实例数	500元/实例/月
运维中心服务扩展包	运维中心规格扩展	<ul style="list-style-type: none"><li>● 日志数</li><li>● span数</li><li>● 短信数</li></ul>	<ul style="list-style-type: none"><li>● 10G日志/天</li><li>● 10万span/天</li><li>● 50条短信/天</li></ul> <p>扩展包费用为：7500元/个/月</p>
运行时引擎专业版	<ul style="list-style-type: none"><li>● 开箱即用工程能力</li><li>● 支持多种开发框架</li><li>● 标准化运行环境</li><li>● 标准化运维能力</li><li>● 多种灰度策略和规则</li></ul> <p><b>说明</b> 运行时引擎必须依赖运维中心使用，订购运行时引擎必须选择运维中心。运行时引擎实例数量必须和运维中心保持一致。</p>	应用实例数	<p>应用实例数量20。</p> <p>超过20个应用实例可单独购买实例增量包。</p> <p><b>说明</b> 实例为应用部署节点，一个虚拟机或者容器化部署一个Pod为一个实例。</p>
运行时引擎实例增量包	运行时引擎应用实例数扩展	应用实例数	400元/实例/月
运营中心专业版	<ul style="list-style-type: none"><li>● 多类型数据源接入</li><li>● 大数据计算引擎</li><li>● 数十种开箱即用指标模板</li><li>● 自定义运营视图</li></ul>	事件量、看板数、指标数	2亿事件量、3个看板、20个指标
运营中心事件量增量包	运营中心事件量扩展	事件量	2500元/10亿事件量/月



套餐包类型	说明	计费项	规格约束
运营中心看板增量包	运营中心看板扩展	看板数	4000元/看板/月
运营中心指标增量包-10个指标	运营中心指标扩展	10个指标	3000元/10个指标/月
运营中心指标增量包-100个指标	运营中心指标扩展	100个指标	30000元/100个指标/月
运营中心指标管理增强包	提供指标管理高级特性能力：基于语义数据模型，指标定义和计算支持高级模式的计算方式，敏捷开发、灵活高效。		15000元/月
运营中心数智化大屏增强包	提供数智化大屏高级特性能力：灵活的在线构建，开放的离线开发，屏幕动态效果，大屏端呈现等。		15000元/月

计费周期

包年/包月的计费周期是根据您购买的时长来确定的（以UTC+8时间为准）。一个计费周期的起点是您开通或续费资源的时间（精确到秒），终点则是到期日的23:59:59。

例如，如果您在2023/03/08 15:50:04购买了一个月的服务，那么其计费周期为：2023/03/08 15:50:04 ~ 2023/04/08 23:59:59。

计费示例（无增量包）

假设您在2023/03/08 15:50:04购买了应用平台运维中心专业版+运行时引擎专业版，购买时长为1个月，并在到期前手动续费1个月，则：

- 第一个计费周期为：2023/03/08 15:50:04 ~ 2023/04/08 23:59:59
- 第二个计费周期为：2023/04/08 23:59:59 ~ 2023/05/08 23:59:59

费用计算过程如下：

- 运维中心专业版费用：10000元 \* 2个月 = 20000元
- 运行时引擎专业版费用：8000元 \* 2个月 = 16000元

综上，2个月的费用 = 20000 + 16000 = 36000元

计费示例（包含增量包）

假设您在2023/03/08 15:50:04购买了应用平台开发中心专业版，购买时长为1个月，同时订购了5人的增量包（开发人员增量包以人数计算）。那么1个月的费用计算过程如下：

- 开发中心专业版费用：10000元 \* 1个月 = 10000元

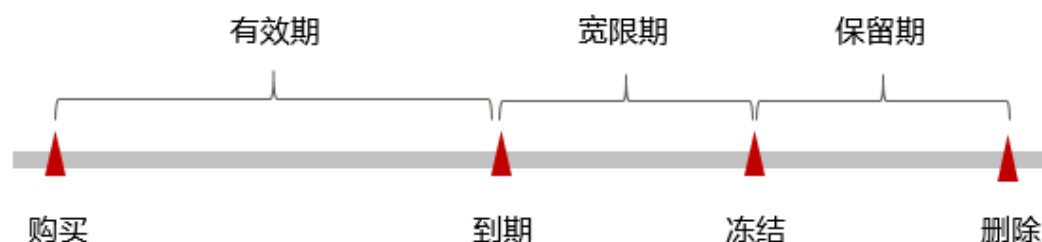
- 增量包费用：200元 \* 5人 \* 1个月 = 1000元

综上，1个月的费用 = 10000 + 1000 = 11000元

## 到期后影响

图2-1描述了包年/包月应用平台服务各个阶段的状态。购买后，在计费周期内资源正常运行，此阶段为有效期；资源到期而未续费时，将陆续进入宽限期和保留期。

图 2-1 包年/包月应用平台服务生命周期



### 到期预警

包年/包月应用平台服务在到期前第7天内，系统将向用户推送到期预警消息。预警消息将通过邮件、短信和站内信的方式通知到华为云账号的创建者。

### 到期后影响

当您的包年/包月应用平台服务到期未续费，首先会进入宽限期，资源状态变为“已过期”。宽限期内您可以正常访问应用平台。

如果您在宽限期内仍未续费应用平台服务，那么就会进入保留期，资源状态变为“已冻结”，您将无法对处于保留期的包年/包月资源执行任何操作。

保留期到期后，若应用平台服务仍未续费，那么服务将无法访问，该资源会自动被删除，且不能再找回，也不能再续费。

### 说明

- 华为云根据[客户等级](#)定义了不同客户的宽限期和保留期时长。
- 关于续费的详细介绍请参见[续费概述](#)。

## 2.3 按需计费

按需计费是一种先使用再付费的计费模式，适用于无需任何预付款或长期承诺的用户。本文将介绍按需计费资源的计费规则。

### 适用场景

按需计费适用于资源需求波动的场景，在使用应用平台的用户创建应用、测试模型服务时需要调用Baichuan2-13B、ChatGLM3-6B模型服务，按需计费模式能大幅降低客户的业务成本。

### 适用计费项

AI原生应用引擎以下计费项支持按需计费。

表 2-3 适用计费项

计费项	说明
大模型微调服务API 在线调用-SFT局部调 优	调用大模型Lora微调服务API时根据实际的输入和输出 tokens总量收费（平台提供算力支持）。

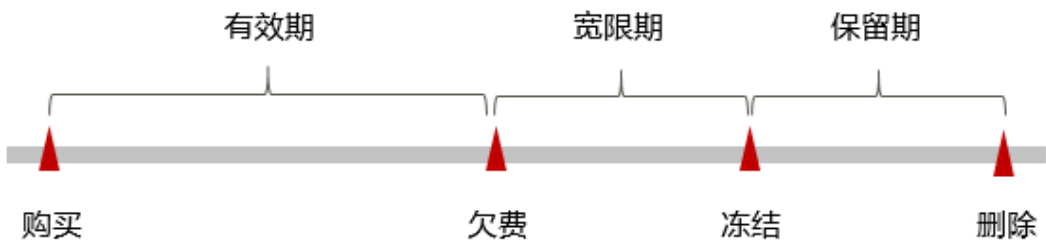
计费周期

API在线调用按需计费按token数量计费，与周期无关。

欠费影响

图2-2描述了按需计费AI原生应用引擎资源各个阶段的状态。购买后，在计费周期内资源正常运行，此阶段为有效期；当您的帐号因按需自动扣费导致欠费后，帐号将变成欠费状态，资源将陆续进入宽限期和保留期。

图 2-2 按需计费 AI 原生应用引擎资源生命周期



欠费预警

系统会在每个计费周期后的一段时间对按需计费资源进行扣费。当您的帐户被扣为负值时，我们将通过邮件、短信和站内信的方式通知到华为云帐号的创建者。

欠费后影响

当您的帐号因按需自动扣费导致欠费后，帐号将变成欠费状态。欠费后，按需资源不会立即停止服务，资源进入宽限期。您需支付按需资源在宽限期内产生的费用，相关费用可在“管理控制台 > 费用中心 > 总览”的“欠费金额”查看，华为云将在您充值时自动扣取欠费金额。

如果您在宽限期内仍未支付欠款，那么就会进入保留期，资源状态变为“已冻结”，您将无法对处于保留期的按需计费资源执行任何操作。

保留期到期后，若您仍未支付帐户欠款，那么推理单元将被释放，数据无法恢复。Baichuan-13B大模型服务API在线调用、bge-large-zh-v1.5 API在线调用、ChatGLM3-6B大模型服务API在线调用、大模型微调服务API在线调用-SFT局部调优将无法再使用。

说明

- 华为云根据[客户等级](#)定义了不同客户的宽限期和保留期时长。
- 关于充值的详细操作请参见[帐户充值](#)。

# 3 计费项

## 计费说明

应用平台服务的计费项如表3-1所示。

表 3-1 应用平台服务计费说明

计费项	说明	规格约束	适用的计费模式
开发中心专业版	<ul style="list-style-type: none"><li>统一应用架构建模</li><li>产品/版本/个人级协作管理</li><li>一站式开发工具链</li><li>研发效能洞察</li><li>海量资产模板</li></ul>	开发中心用户数量50人。 超过50人可单独购买开发人员增量包。	包年/包月
开发人员增量包	开发中心用户数扩展	200元/用户/月	包年/包月
运行时引擎专业版	<ul style="list-style-type: none"><li>开箱即用工程能力</li><li>支持多种开发框架</li><li>标准化运行环境</li><li>标准化运维能力</li><li>多种灰度策略和规则</li></ul> <p><b>说明</b> 运行时引擎必须依赖运维中心使用，订购运行时引擎必须选择运维中心。运行时引擎实例数量必须和运维中心保持一致。</p>	应用实例数量20。 超过20个应用实例可单独购买实例增量包。 <b>说明</b> 实例为应用部署节点，一个虚拟机或者容器化部署一个Pod为一个实例。	包年/包月
运行时引擎实例增量包	运行时引擎应用实例数扩展	400元/实例/月	包年/包月

计费项	说明	规格约束	适用的计费模式
运维中心专业版	<ul style="list-style-type: none"><li>• 视界监控服务</li><li>• 纪实日志服务</li><li>• 追踪者调用链服务</li><li>• 先知告警服务</li><li>• 智维AIOps</li><li>• 蓝军故障演练服务</li></ul> <p><b>说明</b> 运维中心不依赖于运行时引擎，可以只订购运维中心，不订购运行时引擎。</p>	<ul style="list-style-type: none"><li>• 应用实例数量20</li><li>• 100指标/实例</li><li>• 2G/天/实例</li><li>• 1万span/天/实例</li><li>• 5条语音或短信/天/实例</li></ul> <p>超过上述规格可单独购买实例增量包、服务扩展包。</p> <p><b>说明</b> 实例为应用部署节点，一个虚拟机或者容器化部署一个Pod为一个实例。</p>	包年/包月
运维中心实例增量包	运维中心应用实例数扩展	500元/实例/月	包年/包月
运维中心服务扩展包	运维中心规格扩展	<ul style="list-style-type: none"><li>• 10G日志/天</li><li>• 10万span/天</li><li>• 50条短信/天</li></ul> <p>扩展包费用为：7500元/个/月</p>	包年/包月
运营中心专业版	<ul style="list-style-type: none"><li>• 多类型数据源接入</li><li>• 大数据计算引擎</li><li>• 数十种开箱即用指标模板</li><li>• 自定义运营视图</li></ul>	2亿事件量 超过2亿事件量可单独购买增量包。	包年/包月
运营中心增量包	运营中心事件量扩展	2500元/10亿事件量/月	包年/包月

计费示例（无增量包）

假设您在2023/03/08 15:50:04购买了应用平台运维中心专业版+运行时引擎专业版，购买时长为1个月，并在到期前手动续费1个月，则：

- 第一个计费周期为：2023/03/08 15:50:04 ~ 2023/04/08 23:59:59
- 第二个计费周期为：2023/04/08 23:59:59 ~ 2023/05/08 23:59:59

费用计算过程如下：

- 运维中心专业版费用：10000元 \* 2个月 = 20000元
- 运行时引擎专业版费用：8000元 \* 2个月 = 16000元

综上，2个月的费用 = 20000 + 16000 = 36000元

计费示例（包含增量包）

假设您在2023/03/08 15:50:04购买了应用平台开发中心专业版，购买时长为1个月，同时订购了5人的增量包（开发人员增量包以人数计算）。那么1个月的费用计算过程如下：

- 开发中心专业版费用：10000元 \* 1个月 = 10000元
- 增量包费用：200元 \* 5人 \* 1个月 = 1000元

综上，1个月的费用 = 10000 + 1000 = 11000元

3.1 开发中心/运维中心/运行时引擎/运营中心计费项

计费说明

开发中心/运维中心/运行时引擎/运营中心的计费项如表3-2所示。

表 3-2 开发中心/运维中心/运行时引擎/运营中心的计费项说明

套餐包类型	说明	计费项	规格约束	适用的计费模式
开发中心专业版	<ul style="list-style-type: none"><li>• 统一应用架构建模</li><li>• 产品/版本/个人级协作管理</li><li>• 一站式开发工具链</li><li>• 研发效能洞察</li><li>• 海量资产模板</li></ul>	开发中心用户数	开发中心用户数量50人。 超过50人可单独购买开发人员增量包。	包年/包月
开发人员增量包	开发中心用户数扩展	开发中心用户数	200元/用户/月	包年/包月

套餐包类型	说明	计费项	规格约束	适用的计费模式
运维中心专业版	<ul style="list-style-type: none"><li>• 视界监控服务</li><li>• 纪实日志服务</li><li>• 追踪者调用链服务</li><li>• 先知告警服务</li><li>• 智维AIOps</li><li>• 蓝军故障演练服务</li></ul> <p><b>说明</b> 运维中心不依赖于运行时引擎，可以只订购运维中心，不订购运行时引擎。</p>	应用实例数	<ul style="list-style-type: none"><li>• 应用实例数量 20</li><li>• 100指标/实例</li><li>• 2G/天/实例</li><li>• 1万span/天/实例</li><li>• 5条语音或短信/天/实例</li></ul> <p>超过上述规格可单独购买实例增量包、服务扩展包。</p> <p><b>说明</b> 实例为应用部署节点，一个虚拟机或者容器化部署一个Pod为一个实例。</p>	包年/包月
运维中心实例增量包	运维中心应用实例数扩展	应用实例数	500元/实例/月	包年/包月
运行时引擎专业版	<ul style="list-style-type: none"><li>• 开箱即用工程能力</li><li>• 支持多种开发框架</li><li>• 标准化运行环境</li><li>• 标准化运维能力</li><li>• 多种灰度策略和规则</li></ul> <p><b>说明</b> 运行时引擎必须依赖运维中心使用，订购运行时引擎必须选择运维中心。运行时引擎实例数量必须和运维中心保持一致。</p>	<ul style="list-style-type: none"><li>• 日志数</li><li>• span数</li><li>• 短信数</li></ul>	<p>应用实例数量20。</p> <p>超过20个应用实例可单独购买实例增量包。</p> <p><b>说明</b> 实例为应用部署节点，一个虚拟机或者容器化部署一个Pod为一个实例。</p>	包年/包月
运行时引擎实例增量包	运行时引擎应用实例数扩展	应用实例数	400元/实例/月	包年/包月

套餐包类型	说明	计费项	规格约束	适用的计费模式
运维中心服务扩展包	运维中心规格扩展	应用实例数	<ul style="list-style-type: none"><li>10G日志/天</li><li>10万span/天</li><li>50条短信/天</li></ul> 扩展包费用为：7500元/个/月	包年/包月
运营中心专业版	<ul style="list-style-type: none"><li>多类型数据源接入</li><li>大数据计算引擎</li><li>数十种开箱即用指标模板</li><li>自定义运营视图</li></ul>	事件量	2亿事件量 超过2亿事件量可单独购买增量包。	包年/包月
运营中心事件量增量包	运营中心事件量扩展	事件量	2500元/10亿事件量/月	包年/包月
运营中心看板增量包	运营中心看板扩展	看板数	4000元/看板/月	包年/包月
运营中心指标增量包-10个指标	运营中心指标扩展	10个指标	3000元/10个指标/月	包年/包月
运营中心指标增量包-100个指标	运营中心指标扩展	100个指标	30000元/100个指标/月	包年/包月
运营中心指标管理增强包	提供指标管理高级特性能力：基于语义数据模型，指标定义和计算支持高级模式的计算方式，敏捷开发、灵活高效。		15000元/月	包年/包月
运营中心数智化大屏增强包	提供数智化大屏高级特性能力：灵活的在线构建，开放的离线开发，屏幕动态效果，大屏端呈现等。		15000元/月	包年/包月



计费示例

须知

以下示例中出现的费用价格仅供参考，实际价格请参见各服务价格详情。

- 示例（无增量包）**  
假设您在2023/03/08 15:50:04购买了应用平台运维中心专业版+运行时引擎专业版，购买时长为1个月，并在到期前手动续费1个月，则：
  - 第一个计费周期为：2023/03/08 15:50:04 ~ 2023/04/08 23:59:59
  - 第二个计费周期为：2023/04/08 23:59:59 ~ 2023/05/08 23:59:59费用计算过程如下：
  - 运维中心专业版费用：10000元 \* 2个月 = 20000元
  - 运行时引擎专业版费用：8000元 \* 2个月 = 16000元综上，2个月的费用 = 20000 + 16000 = 36000元
- 示例（含增量包）**  
假设您在2023/03/08 15:50:04购买了应用平台开发中心专业版，购买时长为1个月，同时订购了5人的增量包（开发人员增量包以人数计算）。那么1个月的费用计算过程如下
  - 开发中心专业版费用：10000元 \* 1个月 = 10000元
  - 增量包费用：200元 \* 5人 \* 1个月 = 1000元综上，1个月的费用 = 10000 + 1000 = 11000元

3.2 AI 原生应用引擎计费项

计费说明

AI原生应用引擎会根据用户调用预置模型服务时实际输入和输出tokens总量以及创建推理服务（模型部署）所需购买的推理单元收费。具体内容如表3-3所示。

表 3-3 AI 原生应用引擎计费项说明

计费项	计费项说明	适用的计费模式	计费公式
大模型微调服务 API在线调用-SFT 局部调优	调用大模型Lora微调服务API时根据实际的输入和输出tokens总量收费（平台提供算力支持）。	按需计费	千tokens * 每千tokens单价（0.1/K Tokens）

计费项	计费项说明	适用的计费模式	计费公式
推理单元包周期	将基础或者预训练后的模型部署为在线服务时使用的算力单元，通过购买更多推理单元实现服务QPS扩容。	包年/包月	<ul style="list-style-type: none"><li>按月购买：推理单元数 * 月数 * 6000/推理单元/月</li><li>按年购买：推理单元数 * 年数 * 60000/推理单元/年</li></ul>
问答AI服务基础版	包含问答能力，小规格： <ul style="list-style-type: none"><li>5000万 token 大模型调用</li><li>10个用户（应用最大激活态用户量）</li></ul>	包年	50000/年
	包含问答能力，中规格： <ul style="list-style-type: none"><li>1亿 token 大模型调用</li><li>25个用户（应用最大激活态用户量）</li></ul>	包年	100000/年
	包含问答能力，大规格： <ul style="list-style-type: none"><li>2亿 token 大模型调用</li><li>50个用户（应用最大激活态用户量）</li></ul>	包年	200000/年
问答AI服务专业版	包含问答、问数能力，小规格： <ul style="list-style-type: none"><li>5000万 token 大模型调用</li><li>10个用户（应用最大激活态用户量）</li></ul>	包年	60000/年
	包含问答、问数能力，中规格： <ul style="list-style-type: none"><li>1亿 token 大模型调用</li><li>25个用户（应用最大激活态用户量）</li></ul>	包年	120000/年
	包含问答、问数能力，大规格： <ul style="list-style-type: none"><li>2亿 token 大模型调用</li><li>50个用户（应用最大激活态用户量）</li></ul>	包年	240000/年

## 计费示例

### 须知

以下案例中出现的费用价格仅供参考，实际价格请参见各服务价格详情。

#### 示例：使用包年包月的推理单元资源

假设用户于2024年2月5日10:00:00创建了一个包年/包月的推理单元资源，实例个数为2个，购买时长为2个月，单价为6000/实例/月，则此推理单元资源运行期间产生的费用计算如下：

总费用=6000元\*2个实例\*2个月=24000元

# 4 续费

## 4.1 续费概述

### 续费简介

包年/包月应用平台服务到期后会影响应用平台正常运行。如果您想继续使用，需要在指定的时间内为应用平台续费，否则实例等资源会自动释放，数据丢失且不可恢复。

续费操作仅适用于包年/包月资源，按需计费资源不需要续费，只需要保证账户余额充足即可。

应用平台在到期前续费成功，所有资源得以保留，且应用平台的运行不受影响。

### 续费相关的功能

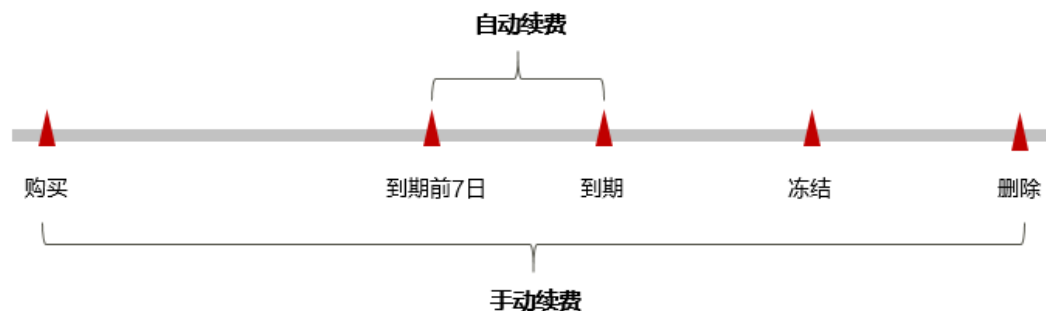
包年/包月应用平台服务续费相关的功能如[表4-1](#)所示。

表 4-1 续费相关的功能

功能	说明
手动续费	包年/包月应用平台服务从购买到被自动删除之前，您可以随时在应用平台控制台为应用平台续费，以延长应用平台的使用时间。
自动续费	开通自动续费后，应用平台服务会在每次到期前自动续费，避免因忘记手动续费而导致资源被自动删除。

在一台包年/包月应用平台生命周期的不同阶段，您可以根据需要选择一种方式进行续费，具体如[图4-1](#)所示。

图 4-1 应用平台生命周期



- 应用平台从购买到到期前，处于正常运行阶段，资源状态为“运行中”。
- 到期后，资源状态变为“已过期”。
- 到期未续费时，应用平台首先会进入宽限期，宽限期到期后仍未续费，资源状态变为“已冻结”。
- 超过宽限期仍未续费将进入保留期，如果保留期内仍未续费，资源将被自动删除。

#### 说明

华为云根据**客户等级**定义了不同客户的宽限期和保留期时长。

在应用平台服务到期前均可开通自动续费，到期前7日凌晨3:00首次尝试自动续费，如果扣款失败，每天凌晨3:00尝试一次，直至应用平台到期或者续费成功。到期前7日自动续费扣款是系统默认配置，您也可以根据需要修改此扣款日。

## 4.2 手动续费

应用平台服务从购买到被自动删除之前，您可以随时通过应用平台控制台的续费入口进入费用中心为其续费，以延长应用平台的使用时间。

### 操作步骤

- 步骤1** 登录**AppStage管理控制台**，默认进入应用平台管理控制台“总览”页。
- 步骤2** 在总览页面的“套餐详情”区域单击“续费”，进入费用中心的“续费管理”页面。
- 步骤3** 自定义查询条件。

可在“手动续费项”、“自动续费项”、“到期不续费项”页签查询全部待续费资源，对资源进行手动续费的操作。

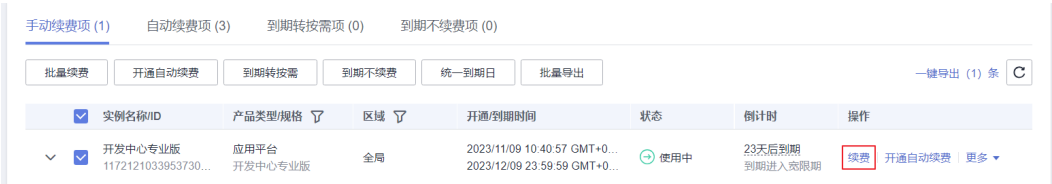
图 4-2 续费管理



所有需手动续费的资源都可归置到“手动续费项”页签，具体操作请参见[如何恢复为手动续费](#)。

步骤4 在列表中找到需要续费的订单，单击操作列的“续费”。

图 4-3 手动续费



步骤5 选择续费时长，判断是否勾选“统一到期日”，将套餐到期时间统一到各个月的某一天（详细介绍请参见[统一包年/包月资源的到期日](#)）。确认配置费用后单击“去支付”。

图 4-4 续费确认



步骤6 进入支付页面，选择支付方式，确认付款，支付订单后即可完成续费。

----结束

## 4.3 自动续费

自动续费可以减少手动续费的管理成本，避免因忘记手动续费而导致应用平台服务无法使用。自动续费的规则如下所述：

- 以应用平台服务的到期日计算第一次自动续费日期和计费周期。
- 应用平台服务自动续费周期以您选择的续费时长为准。例如，您选择了3个月，应用平台即在每次到期前自动续费3个月。
- 在应用平台服务到期前均可开通自动续费，到期前7日凌晨3:00首次尝试自动续费，如果扣款失败，每天凌晨3:00尝试一次，直至应用平台服务到期或者续费成功。
- 开通自动续费后，还可以手动续费应用平台服务。手动续费后，自动续费仍然有效，在新的到期时间前的第7天开始扣款。
- 自动续费的到期前7日自动扣款属于系统默认配置，您也可以根据需要修改此扣款日，如到期前6日、到期前5日等等。

更多关于自动续费的规则介绍请参见[自动续费规则说明](#)。

### 前提条件

应用平台服务包年/包月套餐还未到期。

### 在购买应用平台服务页面开通自动续费

您可以在购买应用平台服务页面开通自动续费，如[图4-5](#)所示。更多购买应用平台服务的介绍，请参见[购买AppStage](#)。

图 4-5 自动续费配置



### 在续费管理页面开通自动续费

**步骤1** 登录管理控制台。

**步骤2** 在主菜单中，选择“费用 > 续费管理”，进入“续费管理”页面。

**步骤3** 自定义查询条件。

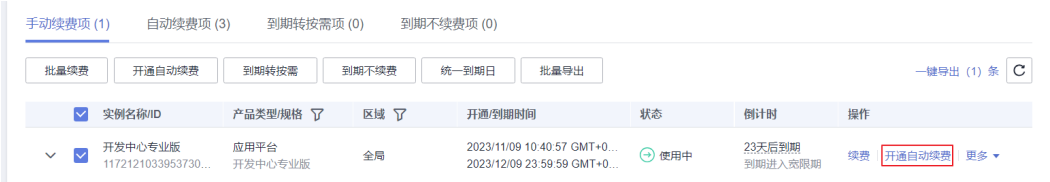
- 可在“自动续费项”页签查询已经开通自动续费的资源。
- 可对“手动续费项”、“到期转按需项”和“到期不续费项”页签的资源开通自动续费。

图 4-6 续费管理



步骤4 选中需要开通自动续费的应用平台服务，单击操作列“开通自动续费”。

图 4-7 续费管理



步骤5 选择续费时长，并根据需要设置自动续费次数，单击“开通”。

----结束



# 5 费用账单

您可以在“费用中心 > 我的套餐”中，查看与应用平台相关的流水和明细账单，以便了解您的消费情况。

## 账单上报周期

包年/包月计费模式的套餐完成支付后，会实时上报一条账单到计费系统进行结算。

按需计费模式的资源按照固定周期上报使用量到计费系统进行结算。按需计费模式产品根据使用量类型的不同，分为按小时、按天、按月三种周期进行结算，具体扣费规则可以参考[按需产品周期结算说明](#)。AI原生应用引擎的按需计费模式按小时进行结算。

按需计费资源的扣费时间可能会滞后于结算周期，例如：按小时结算的专属资源池在8:30删除资源，但是8:00~9:00期间产生的费用，通常会在10:00左右才进行扣费。在“费用中心 > 账单管理 > 流水和明细账单 > 流水账单”中，“消费时间”即按需产品的实际使用时间。

## 查看指定资源的账单

- 步骤1 登录管理控制台。
- 步骤2 在主菜单中，单击“费用”，进入“费用中心”页面。
- 步骤3 在左侧导航栏中，选择“订单管理 > 我的订单”，进入“我的订单”页面。
- 步骤4 在我的订单中，在如下图所示位置单击鼠标左键，复制订单号。

图 5-1 复制订单号




**步骤5** 在左侧导航栏中，选择“账单管理 > 流水和明细账单 > 明细账单”，选择账期，设置统计周期为“明细”，在筛选条件中选择“订单号/交易号”，并输入中**步骤4**复制的订单号，单击图标即可搜索该资源的账单。

图 5-2 查看账单明细



----结束

# 6 欠费说明

用户在使用应用平台服务时，帐户的可用额度小于待结算的账单，即被判定为帐户欠费。欠费后，可能会影响服务资源的正常运行，请及时充值。

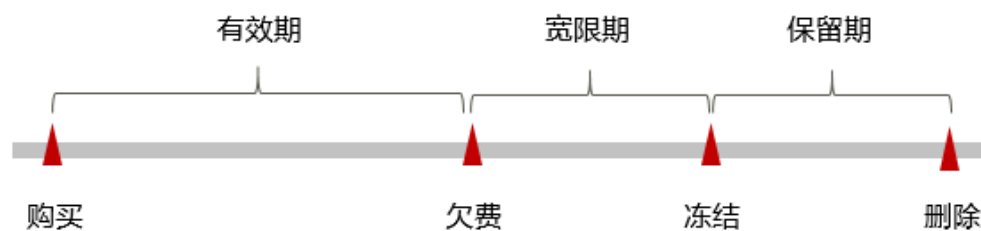
## 欠费原因

- 已购买资源包，但使用量超出资源包额度，进而产生按需费用，同时帐户中的余额不足以抵扣产生的按需费用。请参考[已购买资源包，为什么仍然产生按需计费](#)识别产生按需计费的原因，并重新选择正确的资源包或保证帐户中的余额充足。
- 未购买资源包，在按需计费模式下帐户的余额不足。

## 欠费影响

- 包年/包月  
对于包年/包月的服务资源，用户已经预先支付了资源费用，因此在账户出现欠费的情况下，已有的包年/包月资源仍可正常使用。然而，对于涉及费用的操作，如添加超出开发中心用户数配额的用户、新创建实例、续费订单等，用户将无法正常进行，请按时进行续费。
- 按需计费  
当您的帐号因按需AI原生应用引擎所需服务资源自动扣费导致欠费后，帐号将变成欠费状态。欠费后，按需资源不会立即停止服务，资源进入宽限期。您需支付按需资源在宽限期内产生的费用，相关费用可在管理控制台 > 费用中心 > 总览“欠费金额”查看，华为云将在您充值时自动扣取欠费金额。  
如果您在宽限期内仍未支付欠款，那么就会进入保留期，资源状态变为“已冻结”，您将无法对处于保留期的按需计费资源执行任何操作。  
保留期到期后，若您仍未支付帐户欠款，那么推理单元将被释放，数据无法恢复。Baichuan-13B大模型服务API在线调用、bge-large-zh-v1.5 API在线调用、ChatGLM3-6B大模型服务API在线调用、大模型微调服务API在线调用-SFT局部调优将无法再使用。

图 6-1 按需计费 ECS 资源生命周期



#### 说明

华为云根据[客户等级](#)定义了不同客户的宽限期和保留期时长。

## 避免和处理欠费

欠费后需要及时充值，详细操作请参见[帐户充值](#)。

若存储在对象存储服务 OBS 中的数据不再使用，您可以将他们删除，以免继续扣费。

您可以在“费用中心 > 总览”页面设置“可用额度预警”功能，当可用额度、通用代金券和现金券的总额度低于预警阈值时，系统自动发送短信和邮件提醒。

当产生欠费后，请您及时充值使可用额度大于0。

# 7 停止计费

## 包年/包月资源

- 对于包年/包月计费模式的应用平台服务，用户在购买时会一次性付费，服务将在到期后自动停止使用。
- 如果在计费周期内不再使用包年/包月的应用平台服务，您可以执行退订操作，系统将根据资源是否属于五天无理由退订、是否使用代金券和折扣券等条件返还一定金额到您的账户。
  - 如果您已开启“自动续费”功能，为避免继续产生费用，请在自动续费扣款日（默认为到期前7日）之前关闭自动续费。

## 按需计费资源

对于按需计费模式的资源，例如按需计费的百川开源13B大模型服务API在线调用等，若不再使用这些资源且需停止计费，请删除相应资源。

## 退订包年包月资源以停止计费

- 登录[AppStage管理控制台](#)，默认进入应用平台管理控制台“总览”页。
- 在总览页面的“套餐详情”区域单击“退订”，进入费用中心的“云服务退订”页面。
- 在“退订使用中的资源”页签，单击待退订应用平台套餐后的“退订资源”。

图 7-1 选择退订资源



- 选择退订原因，阅读并确认退订提示，单击“退订”。

图 7-2 退订实例

< 退订使用中的资源

1 您正在退订使用中的资源，请仔细确认资源信息和退款信息。未放入回收站的资源退订后无法恢复，若您要保留资源，仅退订未使用的续费周期，请退订续费周期。

2 属于非五天无理由退订(部分退款)的，需收取手续费和已消费金额，已使用的代金券和折扣券不退还。

实例信息	退订类型	支付信息(¥)	扣减金额(¥)	应退金额(¥)	实际退款(¥)
开发中心专业版 1172121033953730560 产品类型：应用平台 区域：全局	部分退款	10,000.00	-3,258.07	6,741.93	6,741.93
总计		10,000.00	-3,258.07 查看扣费规则	6,741.93	6,741.93

展开退款明细

请选择退订原因

☐ 购买时选错参数

☐ 退订多购买的云服务

☐ 业务测试完毕

☐ 云服务不好用

☐ 不满足业务部署需求

☐ 云服务故障无法修复

☐ 其他

请填写具体原因，100字以内

0/100

实际退款 ¥6,741.93

退还账户余额：¥6,741.93

1 第三方在线支付（如微信、支付宝）退款会返还到华为云账户

2 实际退款金额以账单为准

☐ 我已确认本次退订金额和相关费用，查看退款计算规则

☐ 资源退订后，未放入回收站的资源将立即删除且无法恢复，我已确认数据完成备份或不再使用

退订

退订按需计费资源以停止计费

1. 登录AppStage管理控制台，进入应用平台管理控制台。

2. 在应用平台控制台左侧导航栏选择“AI原生应用引擎”，在AI原生应用引擎详情页面的“我的资源”区域的资源列表中，单击需要停止计费的资源“操作”列的“退订”。

3. 在“退订资源”对话框中，确认要退订的资源，单击“一键输入”自动在输入框填入“退订资源”，如图7-3所示。

图 7-3 退订按需计费的资源

应用平台控制台

模型服务调用统计

预置模型服务调用 API SaaS 自定义模型调用

1 AI原生应用引擎

2 我的资源

3 退订资源

4 退订

4. 单击“确定”，即可退订该按需计费资源。

# 8 计费 FAQ

---

## 如何选择不同的产品？

应用平台AppStage开发中心、运维中心、运营中心可以分开订购，企业根据自己的使用场景按需开通各个中心。运行时引擎不能单独开通，运行时引擎必须依赖于运维中心，即开通运行时引擎必须开通运维中心。

## 运行时引擎开通有何特殊要求？

运行时引擎必须依赖于运维中心，即开通运行时引擎必须开通运维中心。运行时引擎实例数量必须和运维中心保持一致。

## 如何增加告警提示短信容量？

应用平台运维中心提供短信提示功能，默认为5条语音或短信/天/实例，当用户需要更多提示告警容量时，需要购买额外的服务扩展包增加告警发送容量。

# 9 修订记录

发布日期	修订记录
2024-04-16	第三次正式发布。 修改如下章节： <ul style="list-style-type: none"><li>• 包年/包月</li><li>• 按需计费</li><li>• 开发中心/运维中心/运行时引擎/运营中心计费项</li><li>• AI原生应用引擎计费项</li></ul>
2024-02-08	第二次正式发布。 <ul style="list-style-type: none"><li>• 新增如下章节：<ul style="list-style-type: none"><li>- 按需计费</li><li>- AI原生应用引擎计费项</li></ul></li><li>• 修改如下章节：<ul style="list-style-type: none"><li>- 计费模式概述</li><li>- 包年/包月</li><li>- 开发中心/运维中心/运行时引擎/运营中心计费项</li><li>- 续费</li><li>- 欠费说明</li><li>- 停止计费</li></ul></li></ul>
2023-11-25	第一次正式发布。