

# 云性能测试服务

## 产品介绍

文档版本

01

发布日期

2020-08-17



华为技术有限公司



**版权所有 © 华为技术有限公司 2020。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，华为公司对本文档内容不做任何明示或默示的声明或保证。

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。

---

## 目录

---

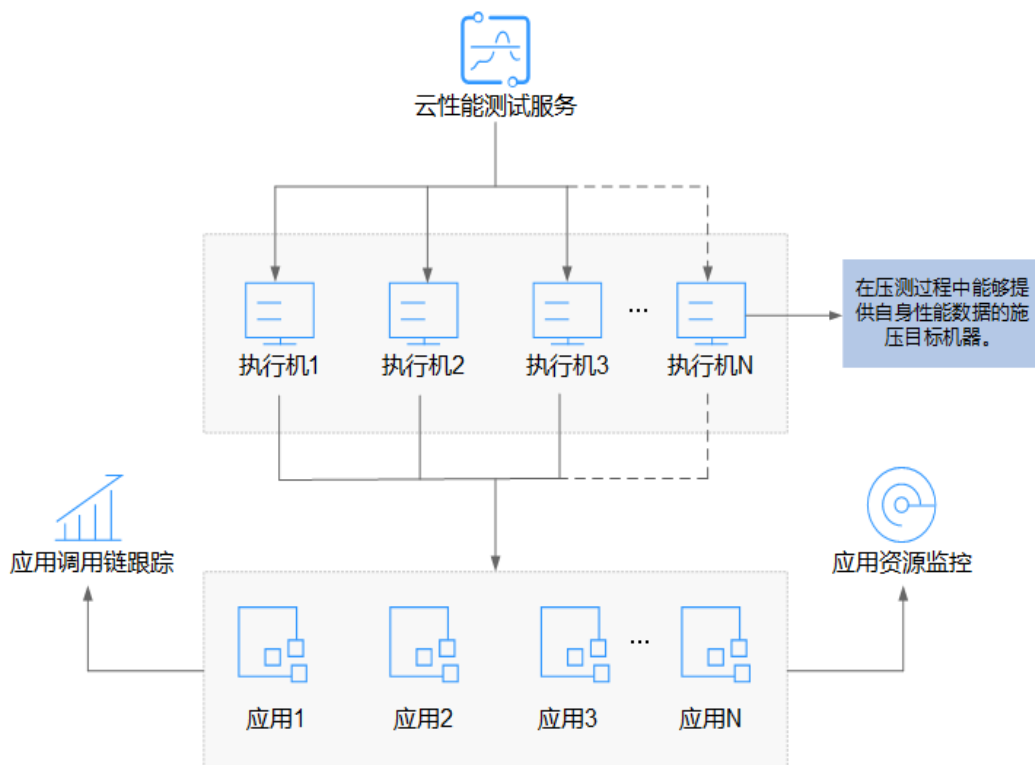
1 什么是云性能测试服务.....	1
2 产品优势.....	4
3 应用场景.....	5
4 约束与限制.....	9
5 基本概念.....	11
6 与其它云服务的关系.....	14
7 计费说明.....	17

# 1 什么是云性能测试服务

随着分布式架构和微服务技术的普及，应用的复杂程度越来越高，在架构解构和性能提升的同时，也带来了生产环境性能问题定位难度高、修复周期长等挑战，因此提前进行性能测试逐渐成为了应用上线前的必选环节。

云性能测试服务（Cloud Performance Test Service，简称CPTS）是一项为基于HTTP/HTTPS/TCP/UDP/WEBSOCKET等协议构建的云应用提供性能测试的服务，具备强大的分布式压测能力。CPTS支持快速模拟大规模并发用户的真实业务高峰场景，可以很好的支持报文内容和时序自定义、多事务组合的复杂场景测试。通过CPTS，我们希望将性能压测本身的工作持续简化，将更多的精力回归到关注业务和性能问题本身，同时降低成本、提升稳定性、优化用户体验，最大程度地帮助企业提升商业价值。

图 1-1 云性能测试服务



## 产品功能

云性能测试服务提供了HTTP/HTTPS/TCP/UDP/WEBSOCKET协议的高并发测试能力，可以支持多协议报文内容、事务、测试任务模型的灵活自定义，可实时、离线查看并发、TPS、响应时延等多个维度的性能统计，同时根据用户对性能测试规模的变化，提供按需的私有测试集群创建、扩缩容等性能测试集群管理能力。

### 多协议高并发性能测试

- 标准HTTP/HTTPS/TCP/UDP/WEBSOCKET报文内容快捷自定义，简单调整即可给不同的被测试应用发送压测流量。

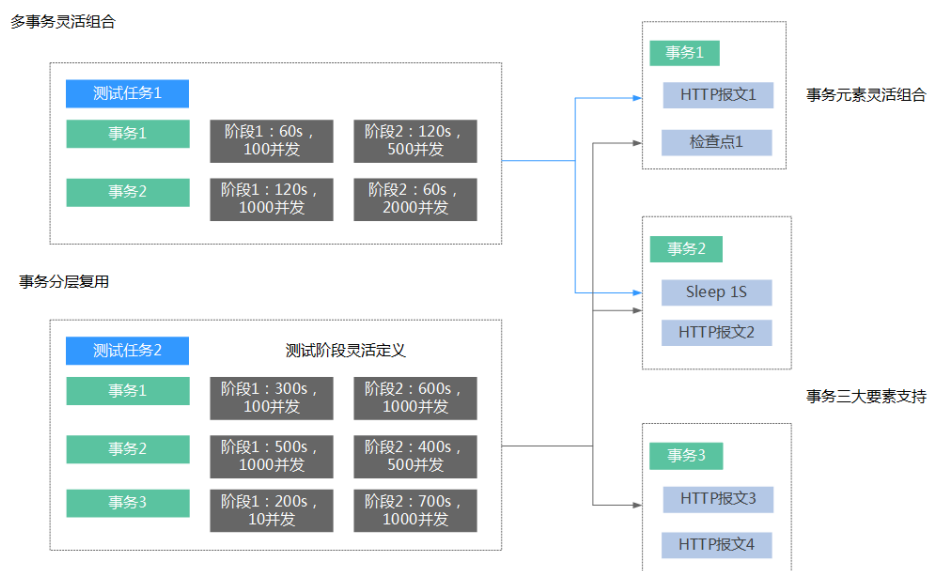
可以根据被测试应用的实际需求，对HTTP/HTTPS/TCP/UDP/WEBSOCKET协议报文的任何字段内容进行自定义，包括HTTP GET\POST方法、URL、Header、Body等字段的设置和编辑。

- 虚拟用户的行为定义，适配不同测试场景。  
通过思考时间对同一个用户的请求设置发送间隔或者在一个事务中定义多个请求报文来设置每个用户每秒内发起的请求数。
- 自定义响应结果校验，更准确的请求成功标准。  
针对每个用户的请求，支持用户配置检查点，在获取到响应报文后针对响应码及头域内容做结果检验，只有条件匹配后才认为是正常响应。

### 测试任务模型自定义，支持复杂场景测试

- 通过多种事务元素与测试任务阶段的灵活组合，可以帮助用户测试在多操作场景并发场景下的应用性能表现。
- 事务可以被多个测试任务复用，针对每个事务可以定义多个测试阶段，并对每个阶段分别定义持续时间和并发用户数或者压测次数，模拟流量波峰波谷的复杂场景。

图 1-2 支持复杂场景测试



### 专业性能测试报告，应用性能表现一目了然

- 提供事务TPS、并发用户、响应时延、访问累计、响应结果校验失败、响应超时等多种细分维度统计能力。

- 提供实时、离线两种类型的测试报告，供用户随时查看和分析测试数据。

**私有压测集群管理，流量租户隔离，用户按需使用**

- 用户按需创建测试集群，实现租户间流量隔离和内/外网压测能力，完成测试后可以随时删除集群。
- 提供测试集群的实时扩容、缩容、升级能力。

# 2 产品优势

云性能测试服务提供一站式性能测试解决方案，帮用户提前识别性能瓶颈。

## 低成本的超高并发模拟

- 能够为用户提供单执行机支持万级并发、整体百万级并发的私有性能测试集群。
- 秒级百万并发能力，模拟瞬间发起大量并发压力，不仅可让企业提前识别高并发场景下应用的性能瓶颈，防止上线后访问过大导致系统奔溃，而且易于操作，极大的缩短了测试时间。
- 支持多任务并发执行，让用户可以同时完成多个应用服务的性能测试，大幅提升测试效率。

## 性能测试灵活快捷，助力应用快速上线

- 协议灵活自定义：支持HTTP/HTTPS测试，适应基于HTTP/HTTPS协议开发的各类应用和微服务接口性能测试；支持TCP/UDP/WebSocket测试，支持字符串负载与16进制码流两种模式，满足各类非HTTP类协议的数据构造。
- 多事务元素与测试任务阶段的灵活组合：提供灵活的数据报文、事务定义能力，结合多事务组合，测试任务波峰波谷，可模拟多用户多个操作的组合场景，轻松应对复杂场景的测试；支持针对每个事务指定时间段定义并发用户数，模拟突发业务流量。

## 性能测试压测资源管理，按需使用

- 云容器引擎（CCE）资源组：用户按需创建测试集群，实现租户间流量隔离和内/外网压测能力，完成测试后可以随时删除集群。同时，提供测试集群的实时扩容、缩容、升级能力。
- 共享（外网）资源组：不需要用户创建，直接使用，调试和小并发压测更方便。

## 快速定位性能瓶颈

- 提供专业性能测试报告，包括事务并发、TPS、吞吐量、响应时延等多维度统计，客观反映用户体验。支持在线和离线报告，方便用户无人值守测试后对测试数据进行分析。
- 无缝对接应用性能管理（APM）、应用运维管理（AOM），通过智能分析功能关联多个监控对象，展示应用资源使用情况、应用调用全链和拓扑关系，实时了解应用的运行状态，快速定位性能瓶颈。

# 3 应用场景

CPTS具备强大的分布式压测能力，应用十分广泛，适合互联网、数字化营销平台、车联网、金融等各行业。

## 电商抢购测试

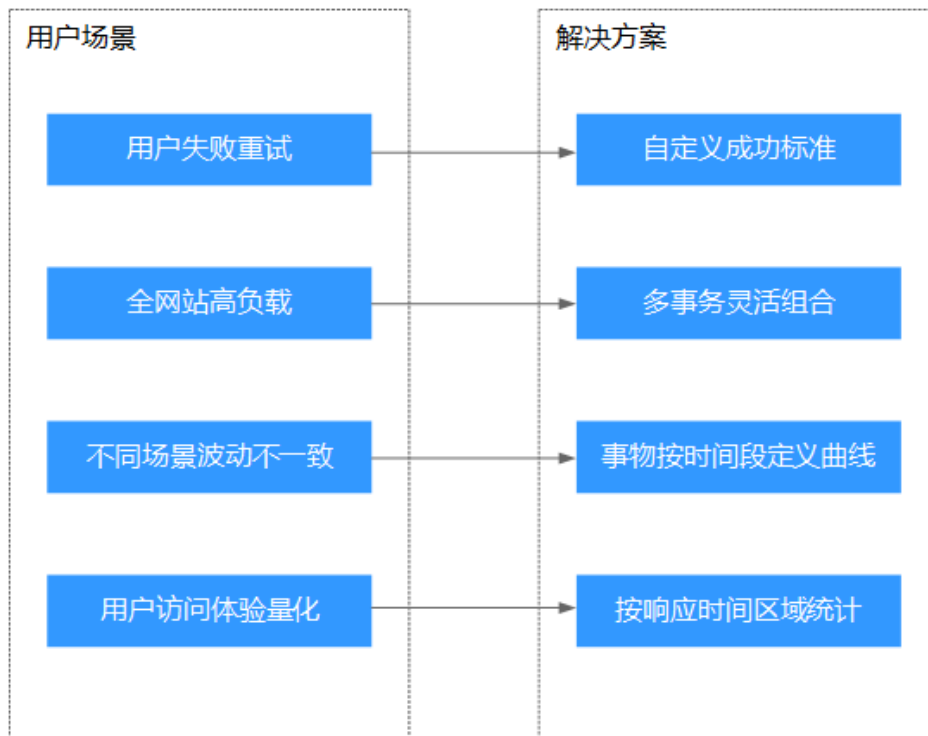
电商抢购已成为当前互联网应用的普遍需求，有并发用户高、突发请求大、失败用户反复重试等特征，如何保证在高负载运行情况下网站的可用性已经成为运维保障的重点。

### 优势

- 真实场景模拟：秒级百万并发能力，瞬间发起大量并发压力，可在一个测试模型里面模拟全网站高负载。
- 专业测试报告：提供按时延响应区间的统计，客观反映用户体验。
- 失败用户重试：多种表达式的自定义结果比对，未正常进入网站的可以重试。



图 3-1 电商抢购测试



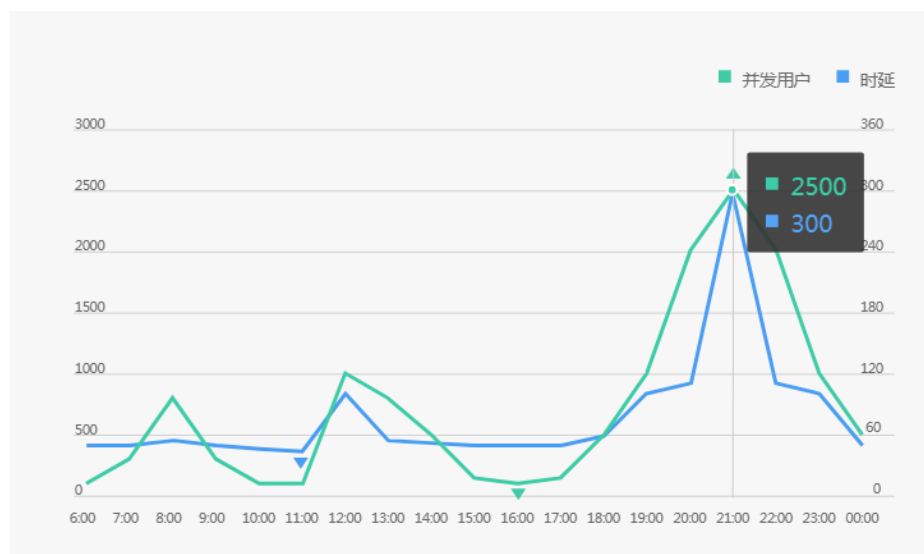
## 游戏高峰测试

游戏行业业务波峰波谷明显，具备弹性伸缩的能力，一方面需要验证弹性伸缩是否可以正常工作，另一方面需要验证在流量突发高峰场景下，时延等关键KPI是否达标。

### 优势

- 多场景组合模拟：通过多事务组合、事务元素多样性、报文自定义功能模拟真实场景。
- 波峰波谷模拟：针对每个单事务根据时间段定义压测曲线，模拟波峰波谷。
- KPI度量：通过自定义响应超时时间，验证高峰场景游戏KPI满足度。

图 3-2 游戏高峰测试



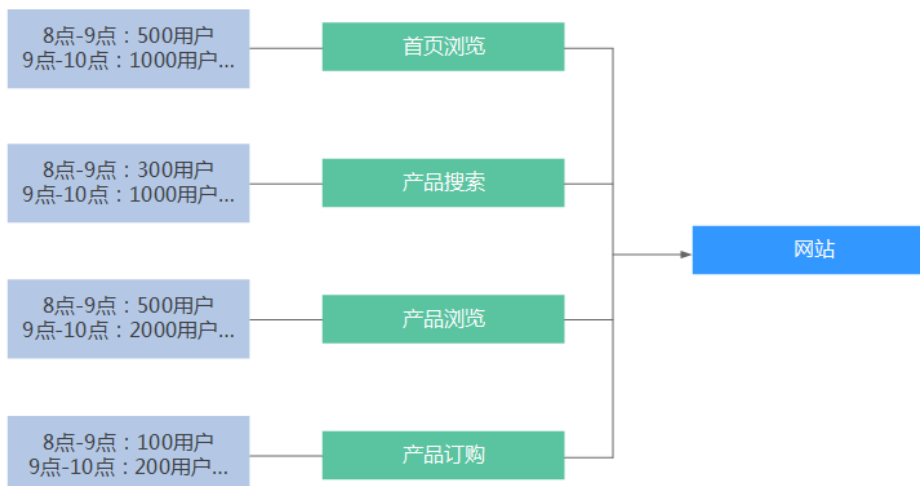
## 复杂场景支持

生产环境往往是复杂多变的，如一个用户访问可能包含多个请求，不同的用户在进行不同的事务操作，用户访问呈现明显的波峰波谷，瞬时并发用户多等状况，因此需要对服务开展性能测试，提前识别性能瓶颈。

### 优势

- 模型灵活定制：支持多事务组合测试，可模拟多用户多个操作的组合场景。
- 突发流量支持：支持针对每个事务指定时间段定义并发用户数，模拟突发业务流量。
- 结果校验：支持多种表达式的自定义结果比对，定制您的事务成功标准。

图 3-3 复杂场景支持



## 应用性能调优

定义性能测试模型，通过云性能测试服务的执行机给被测应用发送模拟流量，利用服务报告查看被测应用的资源监控、调用链情况，了解应用对事物的并发处理能力，方便进行性能优化。

### 优势：

- 灵活扩展：执行机集群按需扩展，支持不同规模的性能测试。
- 一站式解决方案：通过专业的报告提供应用并发能力、响应时延、应用CPU/内存占用、应用的内部各微服务处理时延等关键指标。
- 按需收费：根据性能测试持续时间、并发规模进行收费。

# 4 约束与限制

在使用压测集群前请先通过[云容器引擎](#)创建部署性能测试集群所需的资源。建议根据  
需要压测的并发用户数，在云容器引擎中创建对应规格的节点。

CCE集群至少创建2台节点，1台用于压测集群的管理机（即管理执行机的节点），1台  
用于压测集群的执行机（即在压测过程中能够提供自身性能数据的施压目标机器）。

## 规格推荐

根据需要压测的并发用户数，在云容器引擎中创建对应规格的节点，规格推荐见[表 4-1](#)。

表 4-1 规格推荐

并发用户数	所需规格	数量
0-5000	管理节点：4U8G	1
	执行节点：4U8G	1
5001-10000	管理节点：4U8G	1
	执行节点：8U16G	1
10001-20000	管理节点：4U8G	1
	执行节点：8U16G	2
20001-30000	管理节点：4U8G	1
	执行节点：8U16G	3
30001-40000	管理节点：4U8G	1
	执行节点：8U16G	4
40001-50000	管理节点：4U8G	1
	执行节点：8U16G	5
50001以上	管理节点：4U8G	1

并发用户数	所需规格	数量
	执行节点：8U16G	n 说明 每台8U16G的执行节点可支撑10000并发。

## 配额限制

使用云性能测试服务时，需注意以下配额限制，详情请参见[表4-2](#)，更多配额的详细信息请参见[关于配额](#)。

表 4-2 配额限制

参数	描述信息	默认值
单任务最大并发数	单任务最大支持并发数	100000
实例化资源组数目配额	实例化资源组数目限制	5
事务数目配额	单工程事务数目限制	100
单事务元素数目配额	单事务元素数目限制	40
工程数目配额	租户工程数目限制	100
任务数目配额	单工程任务数目限制	200
共享资源组总并发数配额	共享资源组总并发数限制	1000
共享资源组运行任务数目配额	共享资源组运行任务数目限制	2
共享资源组运行任务时长配额	共享资源组运行任务时长限制	3600
文件变量数目配额	文件变量数目限制	10

# 5 基本概念

## 执行机

在压测过程中能够提供自身性能数据的施压目标机器。

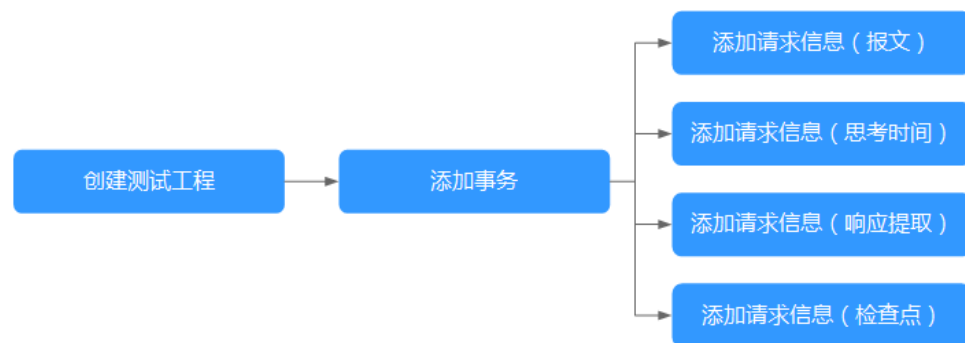
## 管理机

管理执行机的节点。

## 测试工程

CPTS为用户的测试工程提供管理能力，事务、压测任务、测试报告的内容在同一个测试工程内共享复用，您可以为不同的测试项目创建不同的测试工程。

图 5-1 创建测试工程



## 事务

事务是指用户自定义的操作模型，包括思考时间、HTTP/HTTPS/TCP/UDP/WEBSOCKET报文、响应提取和检查点四部分。

## 报文

报文是HTTP应用程序之间发送的数据块。这些数据块以一些文本形式的元信息开头，这些信息描述了报文的内容及含义，后面跟着可选的数据部分。这些报文都是在客户端、服务器和代理之间流动。

## 思考时间

为了更好的模拟用户的行为，需要模拟用户在不同操作之间等待的时间，例如，当用户收到来自服务器的数据时，可能要等待几秒查看数据，然后再做出响应，这种延迟，就称为思考时间。

## 响应提取

如果同一事务中存在多个报文，通过正则表达式把前一个报文的输出提取出来，作后一个报文的输入。

## 检查点

检查点主要是通过自定义校验信息来验证服务端的返回内容是否正确。

## 测试任务

测试任务指基于已定义的测试模型发起一次性能测试的活动。

## 测试报告

测试任务执行完毕后系统会生成测试结果报告，由各项指标TPS、各项指标总量、带宽、响应时延区间比例、TP最大响应时延五部分构成。

## 并发用户数

并发用户数指在同一时刻内，对系统进行业务操作的用户数量，在云性能测试服务中为用户在定义测试任务阶段设置的虚拟用户数。

## TPS

每秒发出的压测请求数量，指云性能测试服务在统计周期内每秒从被测服务器获取到的响应事务实时统计， $TPS = \text{统计周期内的正常返回数} / \text{统计周期}$ 。

## 带宽

记录压测任务运行所消耗的实时带宽变化。上行带宽指从CPTS测试执行机往外发送出去数据的速度；下行带宽指CPTS测试执行机接收到数据的速度。

## 响应时间

指从客户端发一个请求开始计时，到客户端接收到从服务器端返回的响应结果结束所经历的时间。

## 响应超时

在设置的响应超时时间内（默认5s），对应的TCP连接中没有响应数据返回，为响应超时。出现原因一般是被测服务器繁忙、崩溃、网络带宽被占满等。

## 校验失败

从服务器返回的响应报文不符合预期（针对HTTP/HTTPS默认的预期响应码为200），比如服务器返回404、502等。出现原因一般为高并发情况下被测服务无法正常处理导致的，如分布式系统中数据库出现瓶颈、后端应用返回错误等。

## 解析失败

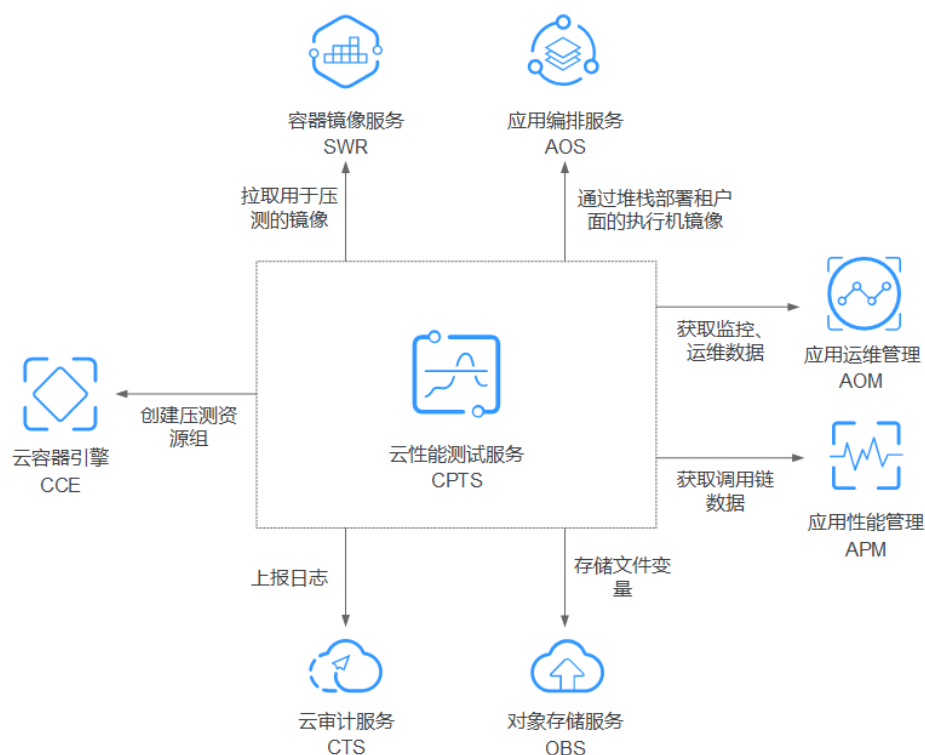
响应报文已全部接收完成，但是部分报文丢失导致整个事务响应不完整，这种情况一般需要考虑网络丢包。



# 6 与其它云服务的关系

云性能测试服务需要与其他云服务协同工作，云性能测试服务和其他云服务的关系如图6-1。

图 6-1 与其它云服务的关系



## 云容器引擎

**云容器引擎**（Cloud Container Engine，简称CCE）提供高度可扩展的、高性能的企业级Kubernetes集群，支持运行Docker容器。借助云容器引擎，您可以在华为云上轻松部署、管理和扩展容器化应用程序。

**CCE和CPTS的交互主要体现在：**通过CCE创建节点，作为CPTS压测资源组中的管理机和执行机。

## 应用运维管理

**应用运维管理**（Application Operations Management，简称AOM）是云上应用的一站式立体化运维管理平台，实时监控用户的应用及相关云资源，采集并关联资源的各项指标、日志及事件等数据共同分析应用健康状态，提供灵活的告警及丰富的数据可视化功能，帮助用户及时发现故障，全面掌握应用、资源及业务的实时运行状况。

AOM作为云上应用的一站式立体化运维管理平台，可以实现对云主机、存储、网络、WEB容器、docker、kubernetes等应用运行环境的深入监控并进行集中统一的可视化管理，能够有效预防问题的产生及快速帮助应用运维人员定位故障，降低运维成本。同时，AOM开放统一API，支撑对接自研监控系统或者报表系统。AOM并非传统监控，它通过应用的角度看业务，满足企业对业务的高效和快速迭代的需求，可帮助企业实现IT对业务的有效支撑，保护、优化IT资产投资，使企业更好的达到其战略目标并实现IT资产调优。

**AOM和CPTS的交互主要体现在：**AOM可提供资源监控，便于进行智能分析。

## 云审计服务

**云审计服务**（Cloud Trace Service，简称CTS），是华为云安全解决方案中专业的日志审计服务，提供对各种云资源操作记录的收集、存储和查询功能，可用于支撑安全分析、合规审计、资源跟踪和问题定位等常见应用场景。

**CTS和CPTS的交互主要体现在：**通过CTS，您可以记录与CPTS相关的操作事件，便于日后的查询、审计和回溯。

## 应用编排服务

**应用编排服务**（Application Orchestration Service，简称AOS）可以帮助您将应用一键式部署到华为云上，简化相关云服务管理操作。AOS通过模板来描述和编排应用及相关云服务，实现自动化部署应用、创建云服务，提供E2E应用全生命周期运维管控能力。

**AOS和CPTS的交互主要体现在：**在CPTS中创建资源组成功后，会在AOS中生成cpts-stack堆栈。

## 应用性能管理

**应用性能管理**（Application Performance Management，简称APM）是实时监控并管理云应用性能和故障的云服务，提供专业的分布式应用性能分析能力，可以帮助运维人员快速解决应用在分布式架构下的问题定位和性能瓶颈等难题，为用户体验保驾护航。

APM作为云应用诊断服务，适用于多种Java框架的应用。它包含了强大的分析工具，通过拓扑图、调用链、事务将应用状态、调用过程、用户对应用进行的操作可视化地展现了出来，以便您快速定位哪些过程出现了问题或者哪里是需要进行改善的性能瓶颈。

**APM和CPTS的交互主要体现在：**应用性能管理平台为压测部署在该平台之上的应用提供调用链跟踪及资源监控服务。在应用性能管理平台的“调用链”界面，可以查看到应用的调用链信息。红色表示异常，黄色表示告警，绿色为正常。

## 对象存储服务

**对象存储服务**（Object Storage Service，简称OBS）是一个基于对象的海量存储服务，为客户提供海量、安全、高可靠、低成本的数据存储能力，包括：创建、修改、删除桶，上传、下载、删除对象等。

**OBS和CPTS的交互主要体现在：**通过OBS，可存储CPTS中的文件变量。

## 容器镜像服务

**容器镜像服务**（Software Repository for Container，简称SWR）是一种支持容器镜像全生命周期管理的服务，提供简单易用、安全可靠的镜像管理功能，帮助用户快速部署容器化服务。

**SWR和CPTS的交互主要体现在：**通过SWR拉取执行集群中用于压测的镜像软件。

# 7 计费说明

## 计费项

云性能测试服务包括两部分费用：所使用资源（弹性云服务器）的费用和使用云性能测试服务的费用。

- 所使用资源主要用于部署性能测试服务的执行集群，由对应的弹性云服务器服务计费，云性能测试服务不再单独收费。  
压测资源组的节点，需提前在云容器引擎中创建，详细步骤请参考[购买节点](#)。资源组中节点由对应的弹性云服务器服务计费。
- 云性能测试服务按压测所消耗的VUM收费，具体计费信息，参见[产品价格详情](#)。

## 计费模式

云性能测试服务支持套餐包，按需和包周期的付费方式。需购买大于压测任务的最大并发数的套餐包，方可使用云性能测试服务执行压测任务，计费时先扣除已购买的有效期内套餐包的额度，超出套餐包部分以按需付费的方式进行结算。

注意：如果您未购买套餐包，将不可使用云性能测试服务。

### 套餐包

套餐包相对于按需付费能够提供更大的折扣，对于长期使用者，推荐该方式。如果您要订购套餐包请单击[这里](#)。

1. 采用预付费模式，一次性支付，即刻生效，暂不支持指定生效日期。
2. 不同的地域之间套餐包不互通，每个地域需分别购买。
3. 套餐包可重复购买，在有效期内叠加使用。
4. 暂不支持续订、退订。套餐包到期后，未使用的VUM会被清零，请留意套餐包即将到期的提醒。您之前使用云性能测试服务产生的数据不受影响，如果您需要再次执行压测任务，则需要购买大于压测任务的最大并发数的套餐包。
5. 购买的套餐包在有效期内，扣费时优先从先过期的套餐包内扣除VUM额度，使用超出套餐包部分，采用按需付费方式进行结算。

例如，您先后购买了三个套餐包，分别为套餐A（50万VUM，最大并发数1万）、套餐B（200万VUM，最大并发数1万）、套餐C（1000万VUM，最大并发数5万）。假如您某一次压测最大并发用户数为8万，您将收到未购买最大并发用户数大于8万套餐包的提示；假如您某一次压测最大并发用户数为5万，最终将抵扣套餐A、套餐B和套餐C中先过期的那个套餐包的VUM额度。

## 按需付费

按需付费比较灵活，可以即开即停。只有运行压测任务后才收费，不执行压测任务不收费。

说明：

对10万以上并发数的压测需要进行工单确认，申请通过后方可进行压测。

1. 登录**工单系统**。
2. 单击左侧导航栏的“工单管理 > 新建工单”。
3. 单击“这里”，填写工单信息，单击“提交”。

图 7-1 新建工单



## 欠费说明

如果您执行压测任务后欠费，将暂停您的性能测试服务。若在保留期（15天）内充值补足欠费后，可查看离线报告数据；若超过保留期未充值，数据将会清除。请注意及时充值，以防数据被清除。