性能测试 PTS

产品简介

产品简介

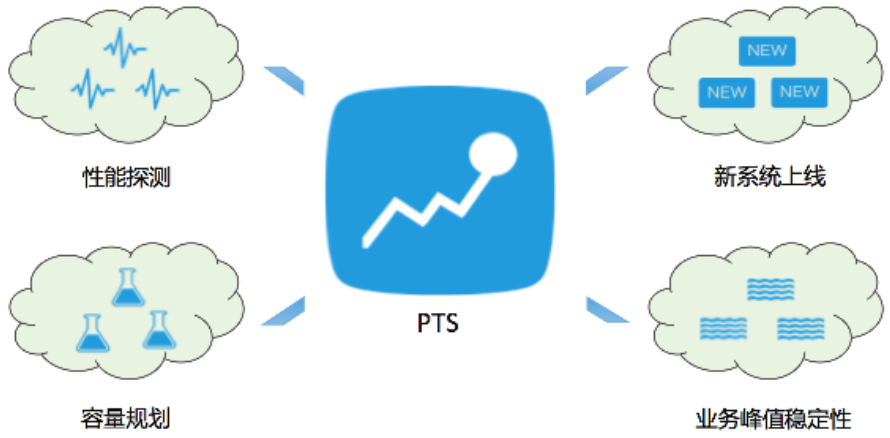
产品概述

性能测试服务（Performance Test Service，简称 PTS）是 Web 化的卓越 SaaS 性能测试平台，具备强大的分布式压测能力，可模拟海量用户的真实业务场景。目前分为铂金版、基础版两个完全不同的版本。

PTS 铂金版是 2016 年 8 月正式发布的全新版本，由阿里巴巴高可用团队倾心打造，是阿里巴巴内部最佳实践的输出。铂金版核心能力基于服务阿里全生态多达 4 年以上的单链路/全链路压测平台。该平台对内除了支持日常的外部流量压测之外，同时支持了大大小小的大促活动，如天猫双 11、双 12 和年货节等等。PTS 铂金版的压力发起来源是遍布全国上百个城市和各运营商的 CDN 节点，相比业界产品的云主机发起更快速，来源更广泛，脉冲能力和流量掌控能力更强。

PTS 铂金版在功能上强调页面可视化编排，目前也在快速迭代中，是接下来 PTS 最重要的版本。相比基础版通过手工编写脚本实现压测的方式，铂金版倡导无需编码的复杂交互式压测。除了上面提及的特性之外，铂金版的 TPS 压测模式、实时调控实时生效的调速能力均领先于业界。

PTS 铂金版目标是将性能压测本身的工作持续简化，使您可以将更多的精力回归到关注业务和性能问题本身。通过 PTS 铂金版可以用最低的人力、资源成本构造最接近真实业务场景的复杂交互式流量，快速衡量系统的业务性能状况，为性能问题定位、容量最佳配比、全链路压测的流量构造提供最好的帮助，进而提升用户体验，促进业务发展，最大程度实现企业的商业价值。



在 PTS 铂金版基础上，我们还可以提供付费的全链路压测咨询、性能及高可用咨询等专家增值服务。

产品功能

压测场景构建

支持有序串行和并行编排压测的 API，参数化上支持数据文件、系统函数、字符串和出参彼此之间的组合，对 cookie 支持非常友好，还有丰富的指令扩展场景的仿真度。调试功能可以便捷地进行复杂场景的数据流向的校验。

相应的资源包配套有极易上手的云端录制，非常便于移动端的请求抓取和到压测场景中的一键导入。

压测流量控制

支持并发和 TPS 模式，分钟内快速启动压测。极低的误差，同时支持自动和纯手动模式，压测流量的调整秒级生效，支持最高千万级的流量瞬时脉冲，多重机制确保压测流量及时停止。

监控和压测报告

陆续丰富中的监控指标，实时监控和报告中包括但不局限于各 API 的并发、TPS、响应时间和采样的日志，请求和响应时间还有不同的细分数据，其他监控能力陆续集成中。

产品优势

稳定可靠

- 阿里巴巴中间件技术部-高可用团队倾心打造，经过内部 5 年以上的全生态沉淀，平台及技术稳定性高；
- 铂金版是基于支持阿里全生态多达 5 年的单链路/全链路压测平台的再加强，内部平台每年支持 10000 次以上的各种大大小小业务压测；
- 铂金版支持的行业广泛，涉及电商、多媒体、金融保险、物流快递、广告营销、社交等等。

功能强大

- 全 SaaS 化形态，无需额外安装和部署；
- 覆盖主流浏览器的录制插件；
- 数据工厂功能，0 编码实现压测的 API/URL 的请求参数格式化；

- 复杂场景的全可视化编排，支持登陆态共享、参数传递、业务断言，同时可扩展的指令功能支持多形态的思考时间、流量蓄洪等；
- 独创的 TPS/并发多压测模式；
- 流量支持动态秒级调整，百万 QPS 亦可瞬时脉冲；
- 强大的报表功能，将压测客户端的实时数据做多维度细分展示和统计，同时自动生成报告供查阅和导出；
- 压测 API/场景均可调试，压测过程提供日志明细查询。

流量真实

- 流量来源于全国上百城市覆盖各运营商（可拓展至海外），真实模拟最终用户的流量来源，相应的报表、数据更接近用户真实体感；
- 施压能力无上限，最高支持千万 TPS 的压测流量。

配套完善

- 除了压测平台之外，可付费增值提供全链路压测解决方案输出，全方位保障站点平稳应对业务峰值。

应用场景

PTS 可以应用但不局限于以下场景：

- 新系统上线，准确探知站点能力，防止系统一上线即被用户流量打垮；
- 峰值业务稳定性，大促活动等峰值业务稳定性考验，保障峰值业务不受损；
- 站点容量规划，对站点进行精细化的容量规划，分布式系统机器资源分配；
- 性能瓶颈探测，探测系统中的性能瓶颈点，进行针对性优化；
- 技术升级验证，大的技术架构升级后进行性能评估，验证新技术场景的站点性能状态。

名词解释

场景

（压测）场景是若干个基于 HTTP/HTTPS 的 URL/API 的组合。URL/API 可能关联了数据文件表示不同用户。不同的 URL/API 表示不同的业务含义（比如登录、加入购物车），最终组合成一个接近用户各种真实行为同时具备一定用户量级的压测模型。

串联链路

指一组压测 API 的有序集合（类似于事务），具有业务含义。压测 API 之间只有在同一个串联链路中才能进行入参和出参关联（运行时数据传递）。两个不同的串联链路之间相互独立，通常不会存在参数的传递依赖（使用数据导出指令的情况除外）。

压测 API

指由用户行为触发的一条端上请求。压测 API 是场景压测中的必需元素，用来定义串联链路中每个阶段 URL 的具体信息。例如，电商网站的登录、查询商品详情、提交订单等，分别对应一次用户行为中的多个请求 API。

出参

从一个压测 API 的应答中截取需要的内容作为出参，供后续的压测 API 作为参数使用。

断言

一般用于标记业务成功与否，从而验证压测请求的响应是否符合预期。有时候响应码是 200 并不代表业务处理成功，有可能需要判断响应体内的内容。在 PTS 的串联链路中如果断言失败，当前请求就不会继续传递到下一个压测 API。另外，在压测实时报表和压测报告中都会相应展现业务成功或者失败的信息。

指令

指令是一种可以改变、控制串联链路中行为和流程的功能组件，可以更真实地模拟业务压测流量。

思考时间

模拟用户在前后两个节点间思考、反应花费的时间，支持多种模式。

集合点

使虚拟用户在集合点处等待，满足条件后一次性释放所有等待的用户，继续后续业务，例如整点秒杀场景。

并发用户数

同时发送压测请求的用户数量。一个用户在压测过程中可能是一个进程或者一个线程。

TPS

每秒发出的压测请求数量。

并发模式

即虚拟用户模式，如果想要摸底业务系统能同时承载的在线用户数，可以通过该模式。

TPS 模式

即吞吐量模式，指每秒固定发出设置的请求数量（TPS）。

响应时间 RT

指从客户端发送一个请求开始，到客户端接收到服务端返回的响应所经历的时间。响应时间由请求发送时间、网络传输时间和服务器处理时间三部分组成。

75% 响应时间

指在整个压测周期内（压测启动到停止的时间内），某个串联链路或者压测 API 的所有采样到的响应时间（固定采样周期）中 75% 的时间在这个值以内。

3xx

这类状态码表示客户端需要采取进一步的操作才能完成请求。通常，这些状态码用来重定向，后续的请求地址（重定向目标）在本次响应的 Location 域中指明。

4xx

这类状态码表示客户端发生了错误，妨碍了服务器的处理。

5xx

这类状态码表示服务器无法完成明显有效的请求。一般代表了服务器在处理请求的过程中有错误或者异常状态发生，也有可能是服务器意识到以当前的软硬件资源无法完成对请求的处理。