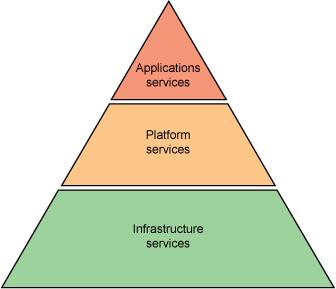
Paas

# 简介

平台即服务（Platform-As-A-Service，PaaS）已经成为业界探讨云计算的热点之一，采用 PaaS 模式构建应用运行平台是实现平台服务的重要实践。

云计算（Cloud Computing）是当前 IT 领域的热点，它的目标之一是通过互联网，使用户更加方便、快捷、灵活地使用各种有质量保障的 IT 资源，这些资源以服务形式提供，终极的云计算环境将使得消费这些服务就像今天使用水、电和煤气等公共基础设施一样便捷。

通常，云计算包括三个主要的层次：基础设施服务（Infrastructure Services），平台服务（Platform Services）和应用服务（Application Services）。如图 1 所示：



其中，平台服务已经成为实践云计算的重点之一，它将应用运行所需的 IT 资源和基础设施以服务的方式提供给用户，包括了中间件服务，信息服务，连通性服务，整合服务和消息服务等多种服务形式。为实现平台服务，业界提出了 “平台即服务（Platform as a Services，以下简称 PaaS）”的交付模式。PaaS 模式，基于互联网提供对应用完整生命周期（包括设计、开发、测试和部署等阶段）的支持，减少了用户在购置和管理应用生命周期内所必须的软硬件以及部署应用 和 IT 基础设施的成本，同时简化了以上工作的复杂度。为了确保高效地交付具备较强灵活性的平台服务，在 PaaS 模式中，平台服务通常基于自动化的技术通过虚拟化的形式交付，在运行时，自动化，自优化等技术也将被广泛应用，以确保实时动态地满足应用生命周期内的各种 功能和非功能需求。

具体来说，搭建传统 IT 基础平台是一个漫长的过程，通常由申请，审计，硬件购买与运输，硬件安装与配置，软件安装与配置等步骤组成。在这个过程中繁复的手工配置工作费时费力，而 且容易产成人为配置错误。同时，平台环境的升级维护也面临人为配置错误频繁产生问题，造成不必要的影响和损失。由于这些原因，搭建完成的应用运行平台，即 使在一定时期内不再需要，也不会被及时释放回收，以供新项目使用。这是造成空闲硬件资源的原因之一。此外，传统基础平台提供的应用运行能力是静态的。然而 在不同时间，应用负载往往是不一样的。为了确保高负载时应用的正常运行，应用运行平台必须能够提供最高运行能力，这就造成了非高峰时的众多空闲硬件资源。

云计算的产生，尤其是平台服务的理念，从产生空闲硬件资源的根本原因入手。建立了快速搭建部署应用运行环境和动态调整应用运行时环境资源这 两个目标。依据虚拟化与自动化技术实现应用运行环境的即时部署以及快速回收，降低了环境搭建时间，避免了手工配置错误，快速重复搭建环境，及时回收资源， 减少了低利用率硬件资源的空置。另一方面，根据应用运行时的需求对应用环境进行动态调整，实现了应用平台的弹性扩展和自优化，减少了非高峰时硬件资源的空 置。

在实际应用中，PaaS 模式的重要应用场景之一是向用户交付一个支撑应用运行的应用运行平台（Applications Running Platform，以下简称 ARP）。基于 PaaS 模式构建应用运行平台，需要满足以下典型需求：

* **提供定义应用需求的接口**：用户可以方便地定义满足应用功能需求所需的组件，同时可以定义应用的非功能性需求，例如可用性，服务水平等；
* **提供基于应用需求快速构建应用运行环境的能力**：能将应用需求映射为物理的 IT 资源和基础设施的具体配置和拓扑结构，并进行快速部署；
* **提供运行时实时动态满足应用需求的能力**：将应用需求映射为运行策略，在运行时，根据系统运行的实际状况，例如负载状况，动态地对物理的 IT 资源和基础设施进行调整，例如调整负载分配或增减计算资源，使 ARP 具备足够的弹性和灵活性，以实时动态地满足应用需求。

针对上述需求，基于 PaaS 模式构建 ARP 的解决方案至少需要包括如下组件：

* **需求与运行策略库：** 向用户提供定义应用需求的界面，保存应用需求，并将应用需求转化为部署及运行策略，提供给平台交付组件和运行时控制组件；
* **软件映像库：** 提供应用运行平台所需的基础软件；
* **ARP硬件资源池：** 基于 Hypervisor，以虚拟化的方式，提供应用运行平台所需的硬件资源；
*  **平台交付组件：** 根据部署或调整策略，通过部署或调整软件及 APR 硬件资源池来创建或调整应用运行平台；
* **运行时控制组件：** 根据运行策略和实际运行状况动态调整 ARP 资源池，并向平台交付组件动态提供调整策略。

