

降本之源

云原生成本管理白皮书



降本之源

云原生成本管理白皮书

编写单位

腾讯云计算（北京）有限责任公司

中国信息通信研究院

作业帮教育科技（北京）有限公司

腾讯云原生 让用云更简单 更有效

目录 CONTENTS

前言	01
一、背景介绍	02
二、云原生成本管理模型	03
2.1 资源利用率现状	04
2.2 云原生成本管理模型	08
2.2.1 成本洞察	08
2.2.2 成本优化	08
2.2.3 成本运营	09
三、最佳实践	10
3.1 成本洞察	10
3.1.1 成本采集及资源追踪	10
3.1.2 资源使用可视化	10
3.1.3 费用可视化	11
3.1.4 成本分配	11
3.1.5 账单管理	13
3.2 成本优化	14
3.2.1 设置合理的资源请求和限制	14
3.2.2 动态调度	17
3.2.3 多维度弹性	19
3.2.4 在离线混部	22
3.2.5 GPU 共享	25
3.2.6 优化矩阵	27
3.3 成本运营	27
3.3.1 建立成本优化团队	28
3.3.2 推动成本意识文化	28
3.3.3 数据驱动成本优化	29
3.3.4 在流程中考虑成本	29
3.3.5 量化成本优化交付的业务价值	29
四、总结	31
五、企业客户降本案例	32

前言

云原生使组织能够在现代云环境（例如公共云、私有云和混合云）中构建和运行可扩展的应用程序，更快地创新并使企业能够更敏捷地对市场作出反应。

凭借弹性计算、自动扩展、计量计费和使用付费模型等功能，云原生计算可帮助组织摆脱昂贵的永远在线的基础架构，并将这些节省用于新功能开发。冗余、容错、松散耦合的服务和云原生架构自动化的自动恢复导致的弹性增加和停机时间减少也可能导致间接成本节约。

使用容器和微服务构建应用程序的一个巧妙之处在于，云原生应用程序使开发人员可以更轻松地访问和重用为早期项目创建的组件。这有几个云原生的优点，可以降低开发成本并制作更好的应用程序：

- **降低开发过程的复杂性：**开发人员可以将更多时间花在项目的细节上，而不是构建通用框架上。它还允许在更短的时间内开发更复杂的应用程序。
- **缩短上市时间：**更快的交付意味着更快的客户更满意，这也意味着抓住时间敏感机会的潜力。
- **简化测试：**经过审查的微服务出现的问题更少，减轻了管道后期的负载。当开发人员知道服务有效时，他们所要做的就是设计兼容性。
- **模块化设计：**模块化设计使外观和功能的标准化的更容易。拥有多个应用程序或服务的企业可以利用这一点来降低客户的学习曲线。

云原生是应用程序开发的未来，具有巨大的业务影响潜力——能够快速有效地将想法转化为生产。云原生已赋能许多其他技术，例如 边缘计算、人工智能、区块链、5G 应用等。

抓住云原生变得特别重要, 你准备好了以下了没:

1. 创建一个更加面向服务的组织。而不是传统的基于功能的结构，围绕特定的服务或能力组织您的团队。
2. 使用现代和最新的架构。云原生意味着使用微服务和反应速度更快的架构类型。
3. 重新组织您的架构以跟上云原生开发的步伐。业务应该能够以与技术人员相同的速度足够快地生成需求。
4. 大多数云原生关键基础设施都是开源的，例如 Kubernetes，你知道如何参与开源社区以了解云原生的最新发展，甚至成为贡献者的一份子吗？

Keith Chan

CNCF 中国区总监兼 Linux 基金会亚太区策略规划总监、腾讯云 TVP

一、背景介绍

数字经济已成为我国经济增长的重要引擎，云计算从部分企业数字化转型的载体，转变为整个经济社会发展的基石与枢纽。万千企业数字化转型提速换挡，对云计算的使用效能提出了更高的要求，云计算迎来了全新的发展机遇。

相较于传统云计算架构，云原生具备更灵活的资源管理、更敏捷的应用迭代、更高效的模块协同以及更稳定的业务保障能力，云原生带动技术架构、应用效能、云化效益的全方位提升，为企业数字化转型提供了有效的实践工具和方法论。

根据中国信息通信研究院调查数据显示，云原生已进入黄金发展期，其核心技术开始在大规模生产环境中深入应用，金融、政府、制造、电信、医疗等行业的云原生用户占比较 2020 年均显著提升，云原生化开始从业内的头部企业逐步下沉到中小企业，从领先企业的尝鲜变为主流企业的必备。同时，云原生技术给企业带来的价值中，提升资源利用率节约成本连续两年排名第一，2021 年已有九成用户认可该项价值，排名前五的另外四项价值分别是：提升弹性效率、提升交付效率、简化运维系统以及便于现有系统的功能扩展。云原生已成为企业基于云实现降本增效的最佳实践。

但随着企业用云程度不断加深，云上支出浪费严重，云原生平台自身的成本治理成为企业上云突出诉求。弹性按需是云原生的资源利用优势，但如果资源配置策略设置不合理可能会导致资源的浪费；云原生资源利用的计量方式如果不够灵活，会使得企业难以准确调用云成本；云边缘、混合多云、云数智融合等模式在帮助提升工作效率、实现应用灵活部署的同时也带来了异构资源管理的新挑战。因此，企业在应用云原生架构之后，需要考虑如何管理、优化和使用云原生服务，如何通过降低云原生成本，进一步提升业务的数字化转型效果。云原生成本管理面临问题难定位、路径难选择、成效难持续三大挑战，资源成本优化是云原生降本的关键路径。

云原生成本管理一是需要掌握成本支出态势，准确定位资源浪费的根源；二是需要选择适合平台架构和业务特点的资源优化策略，并避免因资源优化导致对业务稳定性的影响；三是实现成本优化后还需要保证其持久性。资源成本优化从基于账单优化的成本可视、基于资源优化的成本节约以及基于模式优化的成本运营从三方面帮助降低云原生平台成本。

《云原生成本管理白皮书》将基于中国信息通信研究院与腾讯云对行业发展趋势和企业客户诉求的准确把握、对云原生优化的技术研究与实践经验，提出一套体系化的云原生成本优化方法论和最佳实践路径，结合行业优秀案例，帮助企业改善用云成本充分发挥云原生的效能和价值，为企业的数字化转型提供可靠的保障。

二、云原生成本管理模型

云原生并非一项单纯的技术，更是一种思想，是技术、企业管理方法的集合，云原生追求的是在包括公有云、私有云、混合云等动态环境中构建和运行规模化应用的能力，追求的是业务持续平滑的变更能力。

Kubernetes 是云原生技术栈的核心，是众多云原生项目的粘合剂。Kubernetes 遵循声明式系统原则，将其管控的对象都抽象成标准 API，并通过多种控制器完成云原生平台的高度自动化。比如云用户可以通过定义 Pod 对象，并指定需要运行的容器镜像，以及容器所需的 CPU、内存等计算资源。该对象被提交至 Kubernetes 以后，Kubernetes 会依据用户请求运行容器进程，并按需求确保该应用进程的资源配额。

这使得应用可以以较低成本接入到 Kubernetes 平台中来，并依靠 Kubernetes 自动化机制实现低运维乃至免运维。基于监控平台收集的数据，Kubernetes 可根据应用的实时指标数据，预测应用资源用量，并通过自动化横向或纵向扩缩容能力，及时调整应用的副本数量以及资源用量，及时回收空闲资源，提升资源使用率。

提升资源利用率是云原生技术栈的核心目标之一，资源利用率的提升意味着以更少的计算节点用承载更多的应用实例，极大的降低云用户的资源开销，也契合国家节能减排的政策号召。