

云优化治理白皮书

第一部分：成本优化

云计算开源产业联盟
云优化能力提升计划



2022年5月

版权声明

本白皮书版权属于云计算开源产业联盟，并受法律保护。转载、摘编或利用其它方式使用本调查报告文字或者观点的，应注明“来源：云计算开源产业联盟”或“来源：云优化能力提升计划”。违反上述声明者，本联盟将追究其相关法律责任。

前 言

近年来，随着我国云计算领域的不断发展以及政策的大力推动，企业在云端部署业务系统已经成为了一种趋势，越来越多的企业开始着手关键业务系统上云，用云逐步迈入深水区。然而，深度用云随之而来的成本、性能、稳定性、安全等隐患频现，加速了企业对云的优化和治理需求。

本白皮书是云优化治理系列白皮书的第一部分成本优化。本白皮书首先总结提出企业云成本优化参考框架，覆盖企业云规划、采购、上云、用云和持续运营阶段，提出每个环节可采取的降本增效方式方法。其次，白皮书从云成本优化关键要素和策略方法两个维度进一步展开了云成本优化参考框架的解读。最后，白皮书企业云成本优化需求、成本优化相关技术和工具发展、优化能力参考标准建设四个方面提出云成本优化发展展望。

参与编写单位

中国信息通信研究院、阿里云计算有限公司、腾讯云计算(北京)有限责任公司、天翼云科技有限公司、中国移动通信集团浙江有限公司、优刻得科技股份有限公司、联通(广东)产业互联网有限公司、北京易华录信息技术股份有限公司、中电信数智科技有限公司、富通云腾科技有限公司、广州嘉为科技有限公司、北京佳杰云星数据科技有限公司、国际商业机器(中国)有限公司(IBM)、浩鲸云计算科技股份有限公司、广州青莲网络科技有限公司、北京原森软件科技有限公司、广州汇量信息科技有限公司、北京中软国际信息技术有限公司

主要撰稿人

栗蔚、马飞、李哲、刘学森、叶飞、毕龙飞、刘中巍、冯诗淳、王琳、陆健美、王奇能、王松涛、潘婷、黄雷、杨广贺、王有婷、梁高翔、王健英、郑剑锋、孟凡杰、王孝威、张凌、邢毅彬、罗向军、金天骄、邹文进、龚华兵、曾小龙、安海云、毛钟林、杨玺艳、胡林、穆国华、李枏、姜熠龙、江峰、肖龙强、蔡超

目 录

| | |
|-------------------------------|----|
| 一、云优化背景..... | 1 |
| (一) 企业上云持续深入, 优化治理需求显现..... | 1 |
| (二) 成本优化需求迫切, 落地实施面临挑战..... | 2 |
| 二、云成本优化参考框架..... | 5 |
| 三、云成本优化关键要素..... | 7 |
| (一) 人员是基础, 业务、财务和技术须协同运作..... | 7 |
| (二) 工具是抓手, 监控、分析与调优是必备能力..... | 8 |
| (三) 机制是保障, 成本意识和奖罚机制双轮驱动..... | 9 |
| 四、云成本优化策略方法..... | 10 |
| (一) 规划阶段关注技术选型..... | 10 |
| (二) 采购阶段灵活选择计费方式..... | 12 |
| (三) 上云管云阶段资源有效监管..... | 15 |
| (四) 持续运营阶段成本数据实现长效管理..... | 18 |
| 五、云成本优化发展展望..... | 21 |
| (一) 混合多云将带来云成本优化新需求..... | 21 |
| (二) PaaS 能力建设成为云成本优化新动能..... | 22 |
| (三) 工具一体化和自动化水平将持续提高..... | 23 |
| (四) 优化能力标准体系建设将逐步完善..... | 23 |
| 附录: 云成本优化实践案例..... | 24 |
| (一) 某政府机构私有云成本优化实践..... | 24 |
| (二) 某大型国企混合云架构下的云成本优化实践..... | 26 |
| (三) 某金融公司云成本优化实践..... | 29 |
| (四) 某制造业企业云原生云成本优化实践..... | 32 |
| (五) 某科技公司云成本优化实践..... | 35 |
| (六) 某高校云成本优化实践..... | 37 |
| (七) 某银行云成本优化实践..... | 39 |
| (八) 某科研单位云成本优化实践..... | 42 |

| | | |
|-----|-----------------------|----|
| (九) | 某保险企业云成本优化实践..... | 44 |
| (十) | 某广告公司广告平台云成本优化实践..... | 47 |

一、云优化背景

（一）企业上云持续深入，优化治理需求显现

我国云计算市场保持高速增长，企业上云程度持续加深。2021 年，在全球经济缓慢复苏的背景下，我国云计算市场维持了长期以来的高速增长。据中国信息通信研究院统计，截至 2021 年底，我国云计算市场规模已超 3000 亿元，增速约 45%。云计算作为企业数字化转型的底座，企业上云成为各行业企业数字化转型的必经之路。近年来，企业上云由外围系统过渡到核心系统，企业用云进入深水区。一方面，越来越多的企业基于云原生落地应用，容器、微服务、服务网格等技术成为企业云上业务快速迭代的关键生产力；另一方面，企业开始拥抱多云混合部署模式，中国信息通信研究院的《中国混合云用户调查报告（2021 年）》数据显示，用户平均用云数量达到 4.3 个，企业将应用和数据部署在多云架构上，多云数据治理、多活容灾将成为企业必须考虑的需求。

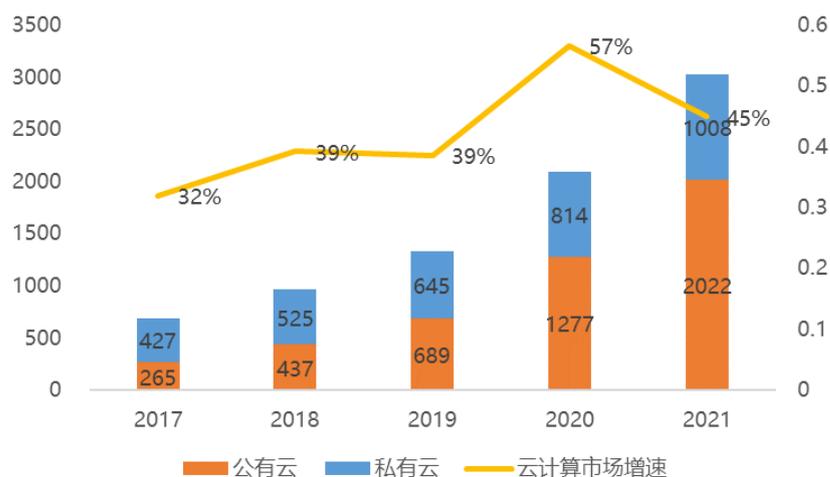


图 1 中国云计算市场规模(单位：亿元)

深度用云随之而来的成本、性能和稳定性等隐患频现，加速企业对云的优化和治理的需求。中国信息通信研究院的《中国云使用优化调查报告》数据显示，75%的企业对当前云使用满意程度较低，其中有47%的企业认为当前云使用效果一般，另外28%的企业表示当前云使用体验较差。企业云优化治理的需求主要表现在以下三个方面，一是用云成本如何优化管理，二是云上业务性能如何调优，三是上云后云平台业务发展可持续性较差。虽然大部分企业已经意识到云优化的重要性，但是企业缺乏云优化相关的能力建设，云优化面临方向不明确、缺乏有效的技术手段、管理流程制度不合理、优化价值难以量化等问题。

（二）成本优化需求迫切，落地实施面临挑战

云成本优化成为企业用云最为迫切的需求。当云计算为企业IT基础设施带来敏捷性和效率提升的同时，企业用云资源配置不合理或配置过渡的现象普遍存在。Flexera 2021 年云状态报告数据显示，除安全之外成本连续两年成为企业最关注的用云挑战，企业上云后平均浪费了30%的云支出，云成本难以控制。《中国云使用优化报告（2021）》数据显示，企业上云后云成本不可控制是企业用云面临的最大挑战，且九成以上用云企业有云成本优化的需求。

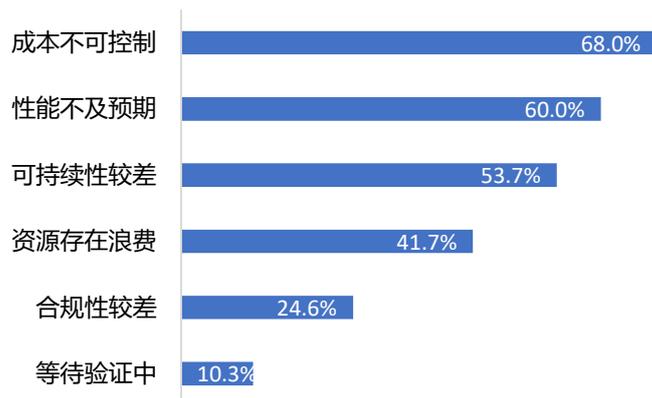


图 2 企业云使用方案不满意情况调查

数据来源:云计算开源产业联盟

云计算重塑企业 IT 消费模型，云成本与传统 IT 固定成本不同，企业 IT 基础设施采购方式从原来的集中式重塑为分布式，且采购决策模式从原来的短期决策重塑为长期运营，企业 IT 成本管理从原来的前置评估重塑为后置监控，且成本类型由原来的固定成本重塑为可变成本，这些变化导致当前云上成本优化管理成为企业用云最具挑战的课题之一，也是企业用云的最大痛点之一。



图 3 云计算重塑企业 IT 消费模型

企业云成本优化实施面临诸多困难和挑战，主要表现在外因和内在两方面：

外在因素，云成本构成复杂。云各种云形态下，企业用云成本构成不同，云成本优化的关注点与实施路径也不尽相同，企业用云的原子模式可以分解为私有云和公有云，在原子模式之上的多云混合场景中，是多种云环境的同时存在，云成本是单一公有云和单一私有云成本的多种组合。每一类部署模式都在不同方面给企业带来成本管理挑战。采用公有云的企业用云成本通常是购买云服务的运营性成本。公有云的资源或服务的种类、计价逻辑和优惠策略的复杂度导致企业对公有云成本管理与优化难度增大。公有云厂商提供的产品数量在数百或上千种，而各类产品的成本计算方式都有其特定规则，同时云厂商又提供了不同的购买方式和优惠策略，在多种复杂度的叠加效应下，云成本的构成变得非常的复杂。

采用私有云的企业用云成本通常包括运营性成本和资本性成本。运营性成本包括人工成本、维修维护成本、电费、公网等费用，这类费用属于数据中心建设后的持续性投入。资本性成本包括机房建设、设备采购、专线铺设、安全设备等费用，这类费用属于前期一次性投入，通常可以使用数年。运营性与资本性成本的组合导致私有云的成本构成极其复杂，在进行成本分摊时，容易出现分摊不均、成本因素缺失等情况。

内在因素，1) 企业缺乏有效的云成本组织管理。企业内部对云成本管理的责任方定义不清晰，若仍采取财务管理制，财务对云的理解是一

大考验。若让 IT 管理部门负责云成本，IT 对成本的敏感性是一大障碍。

2) 云成本优化缺少工具支撑：云成本的一大部分组成在于资源的使用情况，若无可视化工具支撑，企业管理单位很难捕捉各业务的实际使用情况，无法有效进行动态调整。且账单分散在各云服务商处，对于企业统一汇总、分析、优化也带来了不便。**3) 缺少云成本优化的指导方法：**目前国内对云成本的优化处于刚起步阶段，产业相关研究与经验储备不足，业内还未形成通用的优化指导框架，企业只能自行摸索，寻找优化方案。企业云部署模式各异，面临不同成本管理挑战。

二、云成本优化参考框架

当前，企业缺少成本优化相关的标准或建设指南，如何进行优化成为难题。成本优化并不是盲目追求低成本，而是在确保业务价值得到保障的前提下有效提升投入产出比。通常来说云上资源极致的性能、极高的稳定性、极低的成本是云上优化持续追求的三个关键指标，成本优化需要统筹规划多方指标达到最优解，任何时候都不应忽视系统的安全可靠性、可扩展性、性能以及管理效率，这样才能取得最佳的效果。企业上云用云是一个循序渐进的过程，云上成本节约与优化也是一个反复迭代的过程，需要定期分析、持续优化，最终实现最佳运营效果。

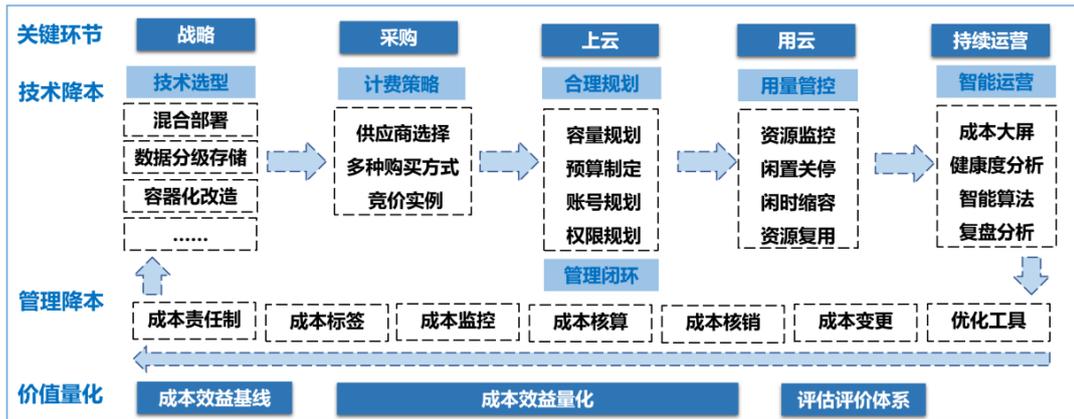


图 4 企业云成本优化参考框架

从企业上云用云历程看，大致可以划分为规划、采购、上云、用云、持续运营等阶段，成本问题贯穿各个阶段，围绕成本问题，每个阶段的关注点又有区别。本白皮书研究提出企业云成本优化参考框架，旨在通过技术手段和管理手段助力企业实现云上成本优化管理，参考框架覆盖企业在云规划、采购、上云、用云和运营各环节的优化策略及实施建议。

规划阶段，企业此阶段要决策云部署及技术选型战略，需要明确用公有云或是自建 IDC 路线，但评估时除了成本外更需要考虑云上技术红利，包括 IAAS 层的更新换代带来的性价比提升，PaaS 层带来的稳定性、性能、研发效率的提升。

采购阶段：企业已确定上云，且在逐步迁移上云中。此阶段要解决如何采购云资源，由于资源类型和计费方式种类复杂，根据业务特性选择适合的资源和付费方式是成本优化最直接的方式。

上云阶段，需要结合企业业务和组织情况提前规划企业云上 IT 治理结构，规划好云上账号、资源、成本、权限、审计等方面的管理体系，

通过治理优先的方式来规划云上的财物权等各关键要素。这些关键问题的背后都与成本密切相关。

用云阶段，企业云计算环境下 IT 资源种类多，且有些资源较隐蔽，容易产生闲置和低效率的情况，此阶段需要关注云资源的使用管控，通过一定的规则和工具识别低效率资源，采取针对性优化措施。

运营阶段，企业应关注云计算成本的长效运作机制，通过云厂工具或者第三方工具进行成本可视化，分析成本与资源的健康度并定期复盘归因，基于历史数据和智能算法对成本预算进行优化变更，使云成本可以持续优化。

三、云成本优化关键要素

人员组织、工具平台和流程制度是企业云成本优化的关键要素。云成本优化并非是通过一系列标准动作后得出的一个终态恒定值，而是一个持续迭代和优化的过程。很多企业做成本优化都是项目制，某个项目效果显著，但是没有长期运作机制。企业需要从人员组织、工具平台和流程制度方面建立云成本优化管理体系关键要素能力。

（一）人员是基础，业务、财务和技术须协同运作

云成本优化团队是企业实施云成本优化的基础。云成本优化不是某一个角色或某一个团队需要做的事情，而是需要多个角色共同参与，打破原有各管一段，各自为战的传统 IT 管理方式，各角色长期协作共同努力以达到成本长期治理的目标。云成本优化团队需要宣传云成本优化意识和文化，推动云成本优化最佳实践，确定企业云成本优化的方向，协

调企业各部门开展成本优化工作。比如，可以定期举行成本优化会议，回顾和复盘成本管理中遇到的一些问题，从而推动持续改进。

企业组建云成本优化团队需要满足以下三个条件，一是人员构成需要跨职能，将云解决方案架构师、业务架构师、运维人员、财务人员、采购人员都纳入到成本管理和长期优化的干系人中来，成本优化团队应由企业各个部门如财务部门、IT 部门、运营部门和业务部门等利益相关者组成的团队。**二是知识体系须完备**，成本优化团队需要具有多学科方法，具备项目管理、数据科学、财务分析和软件/基础设施开发等能力，可以对照成本优化目标（例如资源利用率）来衡量各部门的执行和交付能力。**三是管理层须认可与支持**。支持者即成本优化理念的倡导者，他们会为此团队提供支持，确保按组织确定的优先级开展成本优化活动。优化团队及其支持者会共同确保企业在有效利用云资源并继续创造业务价值。

（二）工具是抓手，监控、分析与调优是必备能力

云成本优化工具是企业实施云成本管理与优化的抓手。云成本优化是一个复杂而耗时的过程，需要将各项能力沉淀到工具。**一方面**，由于云成本的可变性和多云环境的复杂性，云成本优化基于分析结果和优化策略之上并非所有操作都适合人工完成，企业通过工具可以有效提升对云成本的管理及优化水平。**另一方面**，成本优化往往是以项目方式实施，由项目制驱动转为体系化的日常自助优化尤为重要。因此，需要将成本

优化的能力沉淀为工具或平台，构建可度量指标，驱动实际各相关组织自助降本。

企业可以自建或者采购云成本优化工具，工具能力通常需要满足以下三个条件：**一是可以明确资源使用情况并进行调整**，了解资源的使用情况是云成本优化的前提，若无可视化工具支撑，企业很难捕捉各业务的实际使用情况，无法有效进行动态调整。**二是可以关联资源与费用并进行可视化**。费用可视是成本优化的基础，只有看清花费才能针对性的成本优化。成本优化工具需要整合企业各环节云费用支出，支持按组织、部门、业务或产品等维度进行成本展示。**三是工具可以综合分析企业数据与云服务产品进行优化建议**。基于成本优化算法，通过对已购买、未购买的付费模式进行综合对比测算，给出节省计划水位、计费方式等付费模式的优化建议，并具备执行能力。

（三）机制是保障，成本意识和奖罚机制双轮驱动

云成本优化制度是提高企业云成本优化工作成效的抓手。云成本优化相应的长效运营机制在云成本管控中起到关键作用，面对云成本特殊的支出模式和账单结构，企业需要更新一套更加合适的云成本优化流程制度确保优化工作能够在企业内部精确、高效运转。云成本优化流程制度包含三个方面，**一是对企业内部云成本进行权限管控**，包含支出采购、分配修改等。**二是对企业云资源各采购账号进行体系管理**，包含账单核算、托管代付等。**三是对企业各类资源或成本进行统一的配额管理**，包含资源开通、支出审批等。

企业可以通过建立 KPI 制度和响应机制落实云成本优化流程制度。

企业内部云成本日常运维的主要动力和保障来源于云成本优化制度，**一是建立 KPI 制度**，云成本优化作为绩效考核的项目覆盖所有相关部门，根据实际情况合理设置优化目标，如资源闲置率、成本节省额度等，对各部门资源使用进行统一价值量化，按照周期内优化成果进行适当奖罚措施。**二是规划响应机制**，各部门需针对超过预警、预算不足等情况计划应对策略，需通过成本情况进行分析判断、预算调整等方法确保响应机制统一高效运行，对突发情况进行定岗定责，大大缩短内部处理时间。

四、云成本优化策略方法

（一）规划阶段，引入新技术新方案

企业需要关注云计算相关技术最新发展趋势，利用好技术红利带来的成本优化。伴随云计算的发展，新技术、新产品、新工具会不断出现，无论是最新的硬件机型还是最新款的 PAAS 产品亦或是最新的 SaaS 软件，往往都会带来性能或效率上的提升，从而提升企业用云性价比和管理效率。如服务器、数据库的升级，研发效能平台的推出等，企业利用新技术架构或产品类型适配自身业务发展可提升用云效能。

1. 在线/离线混合部署

在线/离线混部是通过在在线业务运行过程中填充离线业务，来提高资源利用率。对于企业同时存在多种业务场景时，可以采取混布方式，可以帮助企业提高云资源的利用率从而降低成本。在线型业务和离线业务具备资源互补的特点，从时间上和对资源的容忍度上可以结合互补。

一方面，在线业务的优先级更高，单机和调度层面会优先保障在线的资源，可能会对离线进行压制和驱逐，另一方面，对于离线任务来说，压制和驱逐对用户是无感的，只需要保证任务顺利完成，有很高的容忍度。

混部概念中在线业务和离线业务的划分如下表所示。

| 在线业务 | 离线业务 |
|-------------|-------------|
| 延时敏感型 | 非延时敏感型 |
| 稳定性要求高 | 稳定性要求不高，可重试 |
| 长期服务 | 短期服务 |
| 搜索类服务 | 大数据、AI 训练等 |
| 日间流量高、夜间流量低 | 无明显变化趋势 |

图 5 在线业务及离线业务对比

通过此方式进行成本优化应注意以下几点内容：

(1) 离线任务不能无限填充，需要保证在线作业不受影响，保证其服务在可接受范围内；

(2) 离线作业要能快速上线下线，当在线作业需要资源的时候，及时出让。

(3) 离线作业运行起来之后，还要保证离线作业的成功率，不能因为频繁出让资源，而导致失败率很高。

2. 冷热数据分级存储

冷热数据分级存储技术通过合理配置业务冷热数据存储资源位置，在保证数据存储效率和应用运行性能的基础上，节省存储资源成本。根据

被访问频率的高低，数据可以被分为热数据、温数据和冷数据，其中，冷数据是指活动不频繁或不会被访问，但仍然需要保留的数据。企业大部分数据受到业务特征、用户行为或监管政策的影响，都会在短时间内变“冷”，这一部分数据集合通常占有所有数据的 60%至 80%，然而并不代表它们失去了价值，企业仍需大量资源来对其进行存储。因此，通过数据分级存储进行成本优化的重点是将冷数据识别并储存在低性能存储资源中，一方面，低性能存储资源更低的价格可以有效地节省企业数据存储成本支出。另一方面，冷数据迁移后热数据在应用中的调取和使用性能将会提升。

3. 采用云原生技术栈

企业采用云原生技术可以构建弹性、松耦合应用系统，通过弹性扩缩容与资源调度机制提升企业用成本效率。一方面，云原生具有多维度弹性能力，容器启停都在秒级，可以高效敏捷的进行扩缩容；且容器共享操作系统内核的设计，可以有效提升资源整体利用率。另一方面，Kubernetes 调度机制可保证资源合理分配，为每个集群找到最适合它的节点，动态调度机制帮助实现了应用层弹性伸缩到资源层弹性伸缩的过渡。通过合理利用 Kubernetes 提供的调度能力，根据业务特性配置合理的调度策略，也能有效提高集群中的资源利用率。

（二）采购阶段，按业务特性组合多种计费方式

1. 选择合理付费方式

根据业务特性选择合适的付费方式是成本优化最直接的方式，从付费层面来看，按照业务场景长期使用的稳定业务资源需求通过选择成本较低的包年包月、预留实例等方式来支撑负载，可以先使用低配置资源，观察评估运行资源负载后升级配置，或将使用率低的资源降低配置或释放。明确业务使用周期的业务且有状态的资源需求可以通过成本相对较高但使用时长选择更灵活的按量付费方式，如临时测试、弹性伸缩选择按量付费资源。按量付费资源开启停机不收费以保留数据并能快速启用。如需长期运行，转预付费计费模式。无状态且可容错的业务可以通过成本非常低廉的抢占式实例来支撑。

| 场景 | 计费模式推荐 |
|----------------|-----------|
| 低优先级负载的业务 | 抢占式实例 |
| 明确业务使用周期的业务 | 按量付费 |
| 需保持稳定性且长期使用的业务 | 包年包月/预留实例 |
| 具有波动特性的业务 | 流量付费 |

图 6 业务场景特点与计费模式选择

此外，付费方式并不是固定不变的，企业应根据业务资源历史运行数据，及时调整资源类型，例如，针对 CPU 利用率一直较高的后付费类（按量计费）云服务器，其连续两个月使用时长产生的费用超过同规格预付费类（包年包月计费）费用，改变其计费模式为预付费。

2. 配置弹性组合购买

企业可以通过组合使用多种类型资源达到节省成本。企业针对云成本的管理与优化需要关注实例的收费模式，云服务商实例定价模式通常有以下三种：

(1)按需：随用随启，根据运行的实例以按小时或按秒的方式计算容量并付费。

(2)预留：有一定的使用承诺（如：1 年，3 年的使用承诺），与按需实例的定价相比，预留实例可提供大幅折扣（通常为按需实例的 60%）。

(3)竞价：极端弹性和廉价的计算资源。它的价格根据供需关系变化，与按量付费实例的相比具有非常明显的价格优势（通常为按需实例的 10%-20%）。

根据业务特性采用资源组合方式是重要的成本优化方式，组合付费是指企业利用云端弹性的优势配置自动伸缩功能，以实现按需购买实例和预留实例的组合使用来应对业务变化资源需求，以获得最佳的成本效益。

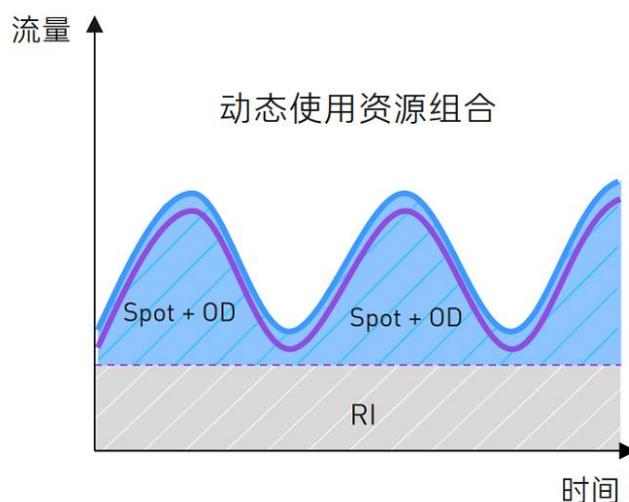


图 7 动态使用资源组合方式

3. 增加竞价实例使用

针对无状态可扩缩的服务采用竞价实例可大幅节省计算资源的成本。

虽然可以以较低的成本购买竞价实例，但是竞价实例存在中断机制，云服务商可根据价格和资源池的存量等情况决定中断运行实例。为了有效的利用竞价实例并保证业务的可用性和稳定性，竞价实例通常应用在无状态可扩缩容的业务上，企业需要对中断进行预判和预处理防止容量损失造成业务宕机，可通过制定竞价回退机制，将受影响的工作负载临时转移到按需节点上，并在竞价节点可访问时将其自动返回到竞价实例来保证容量。

（三）上云管云阶段，关注资源效益完善监管体系

1. 合理规划避免浪费

上云前企业需要对云资源容量进行规划设计，通过管控手段使云成本运营更高效，成本更节省。企业云资源的申请普遍存在过度申请的情况，通常情况下，容量评估主要从常驻资源和弹性资源两个维度进行评估和规划。常驻资源通常承载在线业务、常驻的作业等，这部分的容量规划通常需要根据实际业务场景的水位来预估。对于可靠性要求较高的场景，可以设置峰值水位为常驻容量。而将非预期内的激峰流量和临时任务，交给弹性资源进行补充和供给。企业可通过调研业务部门未来一段时间的业务规划，梳理并列出现适合业务并留有一定冗余量的资源规格和用量。

| 配置 | 适用场景 |
|-------------------------|---|
| 1 核 CPU/2G 内存/1M 公网带宽 | 个人博客、小型网站，Web 应用程序，也可以适用于开发环境、代码存储及测试 |
| 2 核 CPU/4G 内存/3M 公网带宽 | 小程序、APP、论坛、门户类网站，企业运营活动，普通数据处理服务 |
| 4 核 CPU/8G 内存/8M 公网带宽 | APP、视频、购物类网站，ERP 系统，大数据及实时在线或离线分析、深度学习、批量计算、视频编码等 |
| 8 核 CPU/16G 内存/10M 公网带宽 | 游戏开发，高性能科学和工程应用，数据分析、深度学习、批量计算等 |

图 8 常见业务与规格建议

2. 构建资源监控体系

实际运行中的生产系统的底层资源的运行情况才是反应资源利用率的真实的数据。通过构建云资源监控体系，持续监控系统与资源对应的各项指标，一方面，需要监控各类云资源的使用情况，如监控云主机 CPU 使用率、内存使用率等指标，通过时间等维度配置策略。另一方面，需要有效划分闲置资源、低负载资源，如某类云主机从创建后，一直处于关机状态，此类云主机可判定为闲置资源，运维管理员可将云主机进行回收，释放对应的 CPU、内存等资源，避免云资源浪费。

根据监控数据分析出存在优化空间的资源单元。根据持续一段时间周期的监控数据，评估可优化的费用空间，同时对于存在忙时及闲时流量特性的场景，可灵活组合多种计费方式资源以节省成本。例如，监控各业务使用云资源配额的情况，若出现配额闲置，管理员应当减少该租户

部门的容量配额，将配额优先分配给处于瓶颈的业务，减少云平台配额的浪费。此外，在监控的基础上，企业可分析以往数据，有效预测云平台容量未来变化，提前预警提醒企业运维人员。

3. 闲时资源自动缩容

负载访问有明显的周期规律类业务，闲时可以进行资源降配。很多业务的流量具备周期性规律，若是用常备服务器支撑日常峰值流量，在非峰值时段服务器的利用率显示是不足的。对于这一类业务，实际可以根据业务的流量定时扩缩容，可以大幅降低云上成本。



图 9 业务流量周期性规律

业务访问有明显闲时周期的业务通常可分为四类，一是开发、测试类系统，这类系统所用的按量付费资源，非工作时间自动启停。二是社交类业务系统，如微博、抖音等峰值流量出现在中午 1 点和晚上 10 点左右，可能达到日常流量的 1.5 倍甚至更多，可设置阈值自动调配资源。三是证券类业务系统，工作时间内有稳定的访问需求，对于工作时间外例如晚间或者节假日可对资源进行适度降配。四是运行在企业内部的软件系统，例如企业级套装软件 ERP、客户关系管理系统 CRM 和等办公类系统节假日可以适度降低资源配置。

4. 闲置资源及时关停

企业在云上的闲置资源关停通常有两种情况，一是过剩资源关停，当大部分资源的使用率或配额低于了原本规划的范围，且是合理使用时定义为预留过剩。当发现此类情况时，应降低预算配额，合理优化成本。二是遗留资源关停，通常是依赖关系不存在时的遗留资源，例如快照资源，依赖于其他资源生成数据，当其他资源不存在时，快照也无需存在，针对该类资源建议及时停止或删除资源。

（四）持续运营阶段，基于成本数据实现长效管理

1. 统一云成本核算规则

对于有自建数据中心的企业，可以利用成本核算方式来衡量与优化组织的成本使用。制定成本核算模型企业需要明确资源使用定价和成本分摊原则。统一资源定价规则，自建模式没有明显的资源定价水位，需要通过成本估算来制定定价规则，具体包括对包括机房建设、设备采购、机房运维等方面的成本进行统计，将资本性支出和管理支出平摊到云成本中，并在资源实例的定价中体现。混合云场景下可与公有云资源价格进行比较而确定性价比更高的资源配置方案。确定成本分摊原则，需要将数据中心的所有成本完全分摊至使用资源的各个业务部门，则需要将诸如机房建设、运维等费用计入成本，分摊时首先明确成本核销然后以不同维度进行成本的分摊。基于内容定价规则和公有云账单进行例如：单个 CPU/内存的定价需要综合计算物理服务器的购置成本、使用年限、

超配数量、单台运维费用、机房分摊成本、专线接入分摊费用、网络设备分摊费用、运维分摊成本等等。

2. 制定云成本标签体系

从资源和费用维度制定成本标签体系，辅助成本风险容忍度机制对云成本进行监管。监管资源使用状况需要建立资源实例使用指标体系及使用状况告警机制，通过对计算层、存储层、网络层、数据库层实例的相关指标持续跟踪，以判断资源的使用状况。除此之外，通过对个别异常指标的监控告警，可以及时修正资源使用状况，以达到资源使用效益最大化。

监管云成本需要建立云资源成本指标体系及成本风险容忍度机制。有助于评估资源成本的指标体系应包含阶段性支出(月度、季度、年度)、预测支出与实际支出比率、成本增量比率、累计成本、成本趋势等。在确定云资源成本指标体系后，需要协助企业根据业务情况建立风险容忍机制，风险容忍机制的重要因素为容忍度基线及风险响应机制，通过对云资源成本指标建立风险阈值，监测成本风险项，以此及时排查资源使用问题和控制成本。

3. 云成本可视及监管制

用云过程中需要进行成本可视与监管。成本可视，资源的使用情况是影响云成本的主要因素，资源的使用率与成本进行比较后可洞察出可能存在浪费的问题所在。因此，在云成本使用的过程中，企业应具备对

成本使用情况洞察的能力，并通过可视化的工具发现优化点，成本数据可视包括：

(1)账单可视化：企业需具备对云账单可视化的能力，查看账单明细数据，并支持导出进行统一汇总，定期查看实际支出是否与预期符合一致；

(2)账单归属化：企业需根据事前规划，将账单所属部门、业务进行划分，将每份支出落到使用方，从而精细化洞察支出是否与预期符合一致。

成本监管，企业需要设置预算告警、配额监管、时长监管等成本约束手段。预算管理约束，是通过设定成本预算额来约束资源成本的最大值。一般地，会首先制定公司总预算，然后将总预算分解到各部门/成本单元和项目。从而，在实际操作中，不止公司总预算，而且所有成本单元和项目的预算都将作为监管规则存在。如果任一成本单元或项目的预算额度被突破，即使公司总预算未被突破，仍将产生告警。创建预算时，建议按照账号维度、标签维度、财务单元、子账号等多种维度制定预算。

配额和时长监管：一般针对某一种或某一类资源而设置，通常需要根据业务情况定制化。在有些情况下，即使对于同一种资源，也会针对某一配置项的不同取值设定不同的配额和使用时长。

4. 云成本持续闭环运营

云成本优化需要多角色参与长期运营，企业需要建立云上成本优化长期运营机制以及相应指标体系用于分析成本优化效果，监控优化进度。

具体可以通过如下三步来建立企业的成本运营机制及衡量指标体系：

（1）制定成本运营目标

成本优化运营，涉及业务、开发、运维、架构、财务和采购等多团队协作，通过定目标，让项目各成员知道自己和团队的目标是什么，在治理结束之后拿到的什么结果。

（2）基于成本运营数据定制优化方案

成本优化按大类，分运营降本和技术降本。在具体的优化过程中，运营降本和技术降本相互补足，双轮驱动，运营降本过程中碰到的一些难点或通用问题，需要发起专项的技术降本攻克；可复制的技术降本方案，成功经验由点到面放大，需要运营降本持续跟进落地。

（3）定期召开成本会议改进优化策略

利用标签按业务、环境、责任人等维度标记资源，便于日常成本追踪，定期和成本相关方（例如财务、研发等团队）评审预算执行情况，评估优化成果，改进优化策略，建立云成本管理和优化的意识。

五、 云成本优化发展展望

（一）混合多云将带来云成本优化新需求

随着企业上云的逐步普及化，多云、混合云成为了必然趋势，与之带来的是采购模型不一致、计费模型不统一、架构选型复杂化等新的挑战。根据《Flexera2021 年云状态报告》数据显示，92%的企业在 IT 架构

上选择多云战略，企业平均使用 2.6 个公有云+2.7 个私有云，多云、混合云战略是当下大多数企业客户上云的战略选择。在混合多云场景中，公有云、私有云同时存在，加大了云成本优化的难度。

在混合多云环境下的云成本统一治理将成为云成本优化的新挑战。

企业需要统一成本模型和管理流程，减少因环境复杂而带来的管理成本提升，企业未来需要通过工具或手段将公有云、私有云、混合云的账单统一可视、并基于公有云或内部实践的经验制定私有云的计费规则、制定内部结算规则，进而实现在混合多云环境下成本治理的统一化。

企业在分别解决单云场景挑战的同时还需要协调进行多云成本统一管理，一方面，企业面临的是多云账单的统一呈现，协调统一各个异构云账单展示的维度、粒度，通过真实费用账单的进行成本的分析、对比。另一方面，企业各业务系统采用跨云部署的方式部署，需要调整资源在公有云、私有云及专有云的分配比例达到性价比最高。

(二) PaaS 能力建设成为云成本优化新动能

从宏观的视角来看，如果企业上云后基础架构稳固、业务系统稳定、平台安全、云平台支撑业务快速迭代创新，如果达到了上述目标，也是节约成本的一种体现，节约的是创新试错成本，节约的是被攻击后因业务中断导致的经济损失、企业品牌受损的成本。为了实现上述目标，企业上云需要采用合理的安全架构，需要善于使用云平台成熟的 PaaS 能力，比如微服务、中间件、容器、大数据等技术支持企业业务快速创新。所以，从广义上云成本含义不限于资源成本，应以全局的视角看待成本优化。

(三) 工具一体化和自动化水平将持续提高

随着业务复杂度和云复杂度的提升，对云成本的监控、管理、评估和优化难以通过人工的方式实现，需要通过工具和自动化手段来生成更为便捷、更具科学性的优化策略和建议，采用云成本优化工具成为了必然的选择。因为云成本的复杂性和企业用云的灵活性，云成本优化工具仍处于早起发展阶段，当前更多的集中在预算管理、成本分析和账单拆分等几个优化前期过程中，而在成本趋势、优化实现等方面的管理和自动化处理能力还较为薄弱。

未来，云服务商、云管理服务商需要从企业使用的角度，将工具按管理流程进行整合，形成一体化的云成本优化工具，并将积累的大量案例实践和运维经验进行沉淀，通过算法模型进一步实现自动化的成本数据洞察和分析，不断优化模型，使其数据更精准、优化更科学。

(四) 优化能力标准体系建设将逐步完善

云成本优化目前在国内仍处于起步的阶段，随着越来越多的企业认识到云成本优化的重要性，相关的优化标准和评估体系也将逐步建设完善。标准的制定，**一方面**可以指导云服务商、云管理服务商更合理、更全面的提供企业在云成本优化活动中所需要的能力，并通过标准化引导企业在初期可以快速搭建云成本优化管理团队和优化职责。**另一方面**通过评估体系的建设，可以对云服务商、云管理服务商提供方的能力进行有效评估，确保工具的可用性，减少企业因工具选项失败而造成额外的成本增加。并通过评估体系的建设，可以对企业的云成本优化能力与效

果成熟度进行评估，引导企业正确认识自己所处的阶段，采取合理的管理措施。

附录：云成本优化实践案例

（一）某政府机构私有云成本优化实践

1. 企业需求和痛点

某市政府客户，其下辖的各地县、各事业单位的各类内外部业务系统超 600 个，分别部署在私有云、信创云等多种云上；客户在省内各地市有多个数据中心，包含浪潮云、腾讯云、VMware、OpenStack、天翼云等多个厂商的云平台，共计 3000 余台虚拟机，2000TB 存储，每月各类资源支出约 150 万。随着新业务的不断扩大，以及老旧系统的迭代升级，每年依然投入大量的资金在基础设施的建设和维护上。客户的主要痛点主要包括以下几点：

- 1) 资源分布分散、维护难度大；
- 2) 客户对于成本的花费没有具体的数据统计，对于闲置资源没有有效利用，造成资源浪费；
- 3) 对于资源的优化无从下手，资源部署的方式仍然是依据业务需求预估的最高上限申报，无法评估其真实需求并及时调整；
- 4) 无法有效获取私有云的资源消耗数据，整体成本难以统计。

2. 成本优化措施

- 1) 资源评估：对客户的资产规模进行评估，通过客户的 CMDB 及

设备台账搜集整理资产信息，其物理服务器约四分之一的设备已经接近淘汰边缘。在考虑了其数据中心的运营成本及资产性成本后，为客户制定了一套私有云资源价格策略及评估标准。

2) 资源的统一运营及运维：将这些资源信息导入云管平台，由云管平台对多种云资源进行统一资源开通、资源运维、计量、计费，用户只需在一套平台上管理所有云平台的资源，降低了运维复杂度及人员维护的成本，为客户设定各类资源开通、删除、流程共计 12 个；

3) 资源使用控制：通过部门预算、部门配额以及资源池，限定各下级业务部门资源使用的范围、数量及费用；依据资源评估结果，评估上一年度或者月份资源使用的情况，以及今年预计要建设的业务系统规模，设定部门的资源配额、预算及资源池。

4) 私有云资源定价：利用云管平台的成本优化工具对私有云资源进行定价，同时采用统一监控方案，将多云资源的监控信息统一管理，采集所有云资源的性能参数并进行性能分析，通过对近三个月的历史数据的比较分析，推送资源优化建议，并进行资源优化效果模拟。

依据客户资产规模评估结果，为 VMware、OpenStack、浪潮私有云、腾讯私有云的虚拟机实例、数据库实例分别设置定价策略，分别设置单个 CPU、内存、云硬盘价格，同时设置了不同规格的套餐价格。

5) 费率优化及成本预估：依据业务系统对资源的个性化需求，预估应用方案成本规模。结合 Terraform 跨多云资源编排的能力，为业务系统推荐性价比最优的资源落位方案。例如在建设一套财政局内部业务系

统时，可依据资源配置，生成在不同云平台资源配置的价格。

6) 成本分析：按照自然月的方式为客户生成账单，各单位的所有项目的费用均可在平台内统一呈现，不同层级的管理员可以查看的不同的资源范围。市级政府可以查看所有市级单位的资源账单及明细，市级单位可以查看县级单位的资源账单，并结合各业务系统的资源使用率报告，可以评估资源花费是否合理。

7) 资源优化：提供成本分析优化工具，由市级政府的运营管理员统一设置了优化策略，通过监控资源的使用率，当使用率达到某个条件时，识别为冗余资源或闲置资源，例如定义当 CPU 使用率峰值小于 60% 时，识别为资源超配，此时建议降低资源配置目标为 CPU 使用率峰值/0.6。

3. 成本优化效果

系统上线半年后，为客户优化闲置资源 10%，各类业务系统调整降低配置数量达到 30%，客户每月整体的资源费用估算值降低 50 万左右。客户通过云管平台的账单报表可以清晰地分析各个业务系统 IT 基础设施的资源成本，为其调整业务系统决策提供了有效的建议。

(二) 某大型国企混合云架构下的云成本优化实践

1. 企业需求和痛点

国内某大型 IT 企业，在集团内部建立了包括阿里云、Vmware 在内的多个私有云平台，已基本实现了系统全部上云，此外随着研发项目的增多，并且需要经常与客户进行交流、演示，也订购了大量的公有云资

源，包括 800 余台云服务器、40TB 对象存储、42 个弹性 IP 等云资源，每年需要支出大量租用费用。

在采用浩鲸科技的多云管理平台后进行监控分析，发现其在公有云上的云成本浪费较为严重，利用率低于 5% 的云服务器占比 23%，利用率在 5%~10% 区间的云服务器占比 34%，存在公有云资源申请及使用冗余的情况。客户面临的主要痛点包括：

1) 内部研发团队对云资源的需求非常弹性，可能在某一个时间段有大量的使用需求，但是过后很长一段时间不在使用。

2) 对于云成本的支出缺乏整体管控手段，平时依赖各级团队按自己的需求进行把控，年底统一结算时才发现成本支出过多。

3) 对云资源的申请普遍存在过度申请的情况，在实际使用过程中云资源并没有得到利用，造成资源浪费。

2. 成本优化措施

浩鲸科技通过多云管理平台，实现了对私有云与公有云的统一管控，通过其 FinOps 相关能力，实现了申请前阶段的产品与配额控制，申请中阶段的匹配推荐，申请后阶段的精准度量、精细化的成本分析优化与测算，以及 360° 多维视角云成本使用洞察，达到防止云业务过度消费，有效提升云资源利用率，合理降低控制云业务成本的目标。具体实现过程主要包括以下几个方面：

1) 统一管控。利用多云平台实现对企业内部阿里云、Vmware 私有云以及外部租用公有云的统一开通与计费。同时，可以快速接入新的公有云，提供更多的选择。

2) 合理申请。对企业业务特性与云业务场景分析，结合评估历史数据合理配置云成本的预算与云产品的配额，逐级划分到各产品线与团队。完善云资源申请流程，在申请时可根据配额实现多级预警。

3) 最优推荐。及时获取多个云厂商的最新价格与折扣优惠信息包括公有云提供的竞价实例、部分新资源池有更大力度的优化、活动期间包年价更优等策略。针对具体云资源申请业务用途，生成在不同云平台上的费用信息，综合推荐性价比最优的云资源。

4) 资源优化提供成本优化大师工具，由管理员根据业务系统分类与部门分别配置优化策略，通过对云资源使用情况的多维度监控分析，精准识别闲置或未得到充分利用的云资源，并给出合理的优化策略包括降配、或是改变计费模式以减少不必要的成本浪费。提供预估优化前后的比对效果，包括费用降低情况、利用率热力图等，并根据优化策略实现一键优化。例如针对 CPU 利用率一直较高的后付费类（按量计费）云服务器，其连续两个月使用时长产生的费用超过同规格预付费类（包年包月计费）费用，改变其计费模式为预付费；针对演示类、测试类系统，预付费场景下，云服务器 CPU 月平均利用率低于 10%，CPU 利用率中位数低于 25%，降低其规格配置 50~70%（大规格配置资源，降配越多）。

5) 多视角洞察对私有云平台按月生成账单，对公有云平台按月归集账单，并根据云资源的业务归属，按部门维度生成各自的账单信息。客户可从业务系统、部门、云平台等不同视角查看成本分布与消费趋势，一目了然成本结构。

3. 成本优化效果

通过对该企业云成本的优化实践，有效的帮助客户控制了蔓延的云成本，使得成本划分更加准确，成本支出更加清晰，成本使用更加有效。系统上线后，帮助客户识别优化释放闲置资源 8%，优化超配资源 34%，帮助该企业在云成本费用方面每年可减少支出 326 万元，支出降低 28%，取得良好成效。

(三) 某金融公司云成本优化实践

1. 企业需求与痛点

某国内头部基金公司，云主机月持有量 2000+，仅云主机年度支出千万级别，同时拥有阿里云等国内主流云厂商资源，对于多云的资源管理和成本运营一直是其负责的 IT 运维部门和财务部门共同持续关注重点问题，该企业在以往的运营过程中，存在如下需求与痛点：

1) 强稳定性诉求。金融行业客户对业务的稳定性有强诉求，任何资源变动必须以业务稳定性为第一考量；

2) 多云管理难，业务线多，关联资源多，每月将云账单分账至业务线维度需要 2-3 个工作日，消耗大量人力在账单汇集和统筹上；

3) 资源众多，开通灵活，难以厘清是否存在资源闲置、浪费的情况。
在此基础上，多种计费方式组合复杂，人力无法算清最优付费组合策略；

4) 资源管控难，以机器清退为例，下线时缺少业务管控判断，容易出现误删机器的情况。

2. 成本优化措施

围绕成本-性能-稳定性动态最优结合阿里云云资管家平台实现全面资源与成本敏捷运营，进行成本优化之路，具体包括资源治理优化和付费模式优化两部分：

1) 针对闲置、低水位资源的日常管理：系统自动采集资源使用情况。
根据阿里云最佳实践，判断以下条件：

- a. 7 天所有核汇总的 CPU 利用率最大值的 P95 小于 3%
- b. 过去 3 天内平均 CPU 的 P100(所有核的总和) \leq 2%
- c. 七天内出站网络利用率小于 2%

同时符合三项条件的，会被判断为低利用率资源，通过资源运营中心提供预警与建议决策功能，客户会将部分常见性低利用率和闲置规格设定为自动化 workflow 处理。

2) 针对云资源采购付费方式的优化：结合包年包月、SPN、RI 等多种付费模式进行组合优化采购推荐，并且考虑用户资源在时间维度的变化，对客户拥有的资源进行需求分层，识别到当前可能有高风险释放的资源，将这部分资源剔除在计费优化范围外，尽可能保证客户拥有成本节约，降低其超买风险。

同时，采用了需求预测+需求分层来解决客户需求不确定性的问题，具体步骤如下：

- a. 对客户细粒度资源费用进行时序预测，掌握客户未来费用的基本趋势；
- b. 根据客户历史资源释放比例、客户资源保有量的变异系数，根据设定的阈值剔除客户历史开通资源中具有高风险释放的资源；

通过将上述逻辑与算法模型，根据客户过往的消费历史记录通过算法智能推荐给客户多种付费优化结果供客户选择。其中最保守的优化建议预计优化 20% 以上成本，最激进的优化建议预计优化 60% 以上的成本。

3. 成本优化效果

建立了多云环境下资源运营管理中心，达成以下成果：

- 1) 在无任何资源入侵的形式下，仅靠付费方式组合优化，完成总体 21% 以上的成本优化，单账号最高成本优化比例达到 44%；
- 2) 协助客户发现闲置、低利用率资源占总资源数 5% 左右，辅助进行业务决策，提供更广阔的优化空间；
- 3) 固化分摊逻辑，自动化分摊云账单至业务线维度，帮客户看得更轻松，看得更清楚。提升人效 90% 以上，由原 2-3 工作日，优化至 1-2 小时。

(四) 某制造业企业云原生云成本优化实践

1. 企业需求与痛点

作为领先的全球化智能科技公司，TCL 的云上业务系统也具有相当的规模，且先后经历了上云、云原生容器化的过程。随着主要业务从传统 IT 架构迁移上云，TCL 也在进行 IT 企业成本治理的工作。TCL 主要的业务场景主要分为三部分：主要生产业务、压测业务、孵化中的新兴业务。在 IT 企业成本治理视角，不同的业务场景在云原生 IT 成本治理过程中都面对了不同的挑战：

(1) 主要生产业务：

- a) 容量规划难。在从传统 IT 架构迁移云原生后，资源结构发生变化，需要新的容量规划策略。
- b) 不同团队、业务应用的资源分布分散，成本难统计、管理，需要按业务单元细粒度拆分成本。
- c) 业务流量周期性变化，如何更好得进行成本优化。
- d) 多云环境下如何统一进行 IT 成本治理。

(2) 压测业务：

- a) 压测多是临时的任务，闲置资源如何进行成本优化。
- b) 压测环境较生产环境需要大量的机器，如何有效利用已有资源。

(3) 孵化中的新兴业务：

- a) 容量规划难，如何进行资源选型与对新业务应用的容量规划。
- b) Pod 预算难度，增加高可用（某些时候业务量小的应用在节

点调度时会导致业务不可用),在做成本优化的同时会兼顾业务的高可用稳定性。

2. 成本优化措施

1) 全链路压测，容量预估

资源预估是企业基础设施建设的共同难题，也是决定成本规模的先决条件。TCL 在上云期间，采用 PTS 全链路压测，高仿真模拟环境全链路压测，确定系统水位和瓶颈，合理预估资源需求。就遇到生产业务在刚上云阶段难预估、新孵化中的业务难进行容量规划等问题。通过对应用分类，预先确定合适的机型选型以及应用属性配置：首先进行机型选型，根据应用的业务选择适合哪种 CPU/内存比例的机型规格，并在上线时进行动态调整。

2) 混部

TCL 存在大量业务的应用有错峰现象，以及压测场景对资源利用有较高的要求。根据业务的流量波峰波谷、资源使用特性，使用不同的 QoS 服务质量等级策略，如对服务质量要求高的应用选择独占应用属性配置，对服务质量要求不高的应用可与他业务波峰波谷交错的应用共同混部在同一集群节点中。

3) 弹性扩缩

TCL 大量生产应用会根据业务高峰低谷波动，通过设置弹性策略在业务波峰时扩容应用容器副本，以及根据业务波峰自动扩缩节点，在业务波谷时进行缩容，减少波谷时的资源成本开销。

4) 闲置资源回收，动态资源交付

TCL 有大量业务压测场景，压测作业发生后会产生大量闲置资源，TCL 的工程师团队采用动态资源交付的方式，使用阿里云资源按量付费等计费策略，资源即开即用，有效杜绝闲置资源的浪费多租 SaaS 化业务，按业务单元细粒度拆分成本。TCL 的大量业务是采用多租 SaaS 化方式部署在同一集群，业务单元间会相互共享资源，这部分成本难统计的问题，TCL 的工程师团队首先采用阿里云账号体系来拆分粗粒度的静态资源账单，各个业务单元对应到子账号体系，既方便单独管控，也方便进行静态资源的账单拆分，同时，TCL 也参考 ACK 成本分析的集群动态资源账单分析，通过监控、智能推荐等方式调整集群动态资源的配额，实现资源利用率的提升；通过弹性伸缩、动态资源交付等方式，实现资源成本的降低。降本增效的同时，也会大大提升进行 IT 成本治理工作的效率。

5) 多云场景下，云平台标准的成本分析被集成能力

TCL 的大量业务也会在多云场景上运行，面对多云环境下的业务成本难管理的情况，阿里云提供标准的成本分析被集成能力，TCL 的工程师团队使用阿里云平台标准的成本分析统计接口，实现上层业务在多云环境下的成本、资源的统一管控。

3. 成本优化效果

TCL 主生产业务基础设施从传统 IT 架构到上云、再到云原生改造，在这个云原生化过程中，TCL 的业务量也翻了数倍。单个子业务的高

峰期可达到 300+核，成本优化率可达 15%，在进行了混部、业务高峰低谷的弹性伸缩等优化后，优化效果可达高峰期的 30+%的使用率。

(五) 某科技公司云成本优化实践

1. 企业需求和痛点

浙江某集团公司，是一家高新技术科技企业。该公司致力于云渲染、云设计等技术的研发，其开发的装修设计平台软件“KJL”可以实现“所见即所得”的全景VR设计装修新模式（5分钟生成装修方案，10秒生成效果图，一键生成VR方案）。“KJL”平台拥有覆盖全国 90%的户型库，超 300 万室内设计师和超 1000 万业主用户，在国内装修设计平台领域属于领先水平。“KJL”平台所有的计算、存储、渲染都是在云端实现，有的庞大的云资源需求。

目前该公司在浙江省内移动云上订购了 200 台云服务器、100TB对象存储、RDS数据库、20 个 50M弹性IP、云安全等较多资源，目前每年需支付的云成本费用约为 210 万元。

随着企业用云的逐渐深入，浙江移动云MSP专家团队在客户授权背景下对其云上资源利用率进行梳理，发现虚机利用率仅有 32%，容器化利用率仅有 34%，且注意到客户夜间运行离线业务，资源利用率稍微高些，白天资源利用率较低。同时随着调研的不断深入总结出客户上云后出现以下痛点：

①业务增长迅速，业务间有效隔离需要增强；②不同业务独占资源池，出现资源碎片问题；③某些应用独占，资源利用率低且无法腾挪、

复用；④应用服务化需要进行技术改造。

2. 成本优化措施

该企业具体成本优化实现过程如下：

1) 关闭未使用的资源

在企业与MSP团队的配合下，确定云上空闲虚拟机资源、存储等其它资源，并给出关闭资源能够节省的费用账单。

2) 调整使用率较低的资源配置

MSP团队利用自研的优云宝工具查找利用率地下的资源，并通过重新配置或合并这些资源来降低开支。

3) 配置自动缩放

动态分配合取消分配资源来满足性能需求，实现成本节约。

4) 制定预算在合理范围内

借助浙江移动自研的上云宝、优云宝成本管理工具，订阅成本管理服务，同时监视客户云上开支。

3. 成本优化效果

对客户进行云成本优化，通过优化资源碎片、在线/离线混合部署、自动扩缩容等手段，使得该企业客户成本支出方面整体运营成本从原来每年的 210 万元下降到了 153.3 万元，降低了 27%。效率方面整体资源的平均利用率从 33%提升到了 43%。

（六）某高校云成本优化实践

1. 企业需求和痛点

上海某 985 院校以教育数字化转型为核心，搭建一系列教学平台、科研平台、信息共享平台、模型训练平台、竞赛平台等，通过教学数据的积累，用数据驱动的方法去优化教学过程。但在云平台搭建过程中，仍面临了不少成本运营挑战：

1) 规划阶段：由于转型是在动态发展的过程中，无法精准估算后期发展过程中究竟需要配置多少台机器，而以往高校以本地化架构建设为主，若采用一次性采购模式，很容易出现储备不足或者成本浪费的情况；

2) 使用阶段：高校一般采用预算制，预算由校级单位下发给各学院，各学院再分配至各实验课题组，整个成本管理的过程均围绕组织架构进行开展，但公有云多为预付费模式，且混合架构下数据分散无法形成统一监测，成本分摊。

3) 优化阶段：混合架构下，数据分散在各云平台中，若要分析不同课题组资源使用情况为后续发展做优化，工作量大且难度高。

2. 成本优化措施

结合高校业务的特点与客户提出的痛点，优刻得从“上云”、“用云”、“管云”三个阶段出发，以服务咨询结合平台工具的方式，通过云计算技术帮助高校提升教学质量，推动云化转型的同时，减轻高校一次性成本投入，并不断提升资源利用率。

1) 事前上云规划：优刻得与高校老师进行深入交流，了解其教育数

数字化转型的建设框架及使用场景后，我们建议高校将重要的科研数据及师生信息存放在本地数据中心中，对数据量进行三年规划后进行一次性成本投入。而用于学生学习用的教学、训练平台等，因其还处于发展阶段，用量无法准确预估，且需要较为稳定可持续性的服务保障，建议高校部署在公有云平台，利用公有云资源池化的调度能力和 PaaS 组件的平台化能力，随取随用，降低高校的资金投入成本的同时提升科研效率。并通过优刻得 UK8S 部署微服务，使业务模块变成一块块灵活可动的“积木”，随着业务规模的增加，积木变多了，但组合方式依然灵活可变，成本也有效可控。

2) 事中用云监管：考虑到高校采用的是混合架构，资源使用情况分散在各云平台上，在用云的过程中，优刻得为高校部署了自研的 UCMP 多云运营管理平台，从“组织管理”、“预算管理”、“配额管理”、“成本可视”等方面入手。并根据资源的 CPU 使用率和内存使用率等性能指标设定阈值（<5%~10% 区间内）发现闲置、资源浪费的现象，及时干预，对成本使用过程进行持续监管。

3) 事后成本运营：结合优刻得自身公有云运营经验，我们建议老师定期对账单进行责任归属，与规划的预算进行对比，分析当前阶段费用的使用情况是否与预算有较大偏差。并结合 UCMP 的工具手段，自动化生成优化报表，从资源使用率、资源配置分布、成本预测等方面为成本运营提供可视化数据支撑。

3. 成本优化效果

经过三个月的业务梳理与平台验收上线，运行了半年的时间后，学院在多云成本运营方面得到了以下收益：

1) 采购模式优化：由原来的传统本地部署方式转为混合云化部署，减少由于一次性投入而造成的资源成本浪费与人力维护成本。且架构更灵活，更符合教学数字化发展特性。

2) 资源合理化效率提升：部署工具后，对费用账单、资源使用率进行可视化监管后，发现学生在使用过程中对已完成科研的资源未进行及时的回收处理，且存在配置大于实际所需，从而导致资源浪费的情况较多。通过对资源进行及时分析与处理，最终因资源使用不合理造成的成本浪费约减少 20% 以上，资源合理使用率提升 10% 以上。

3) 降低成本运营难度：通过优刻得云管理经验的传递及工具的辅助支撑，简化老师在云化转型中面临复杂性，由执行思维转换为运营思维，通过建立完善的云管理制度，释放老师更多的精力专注于科研上，降低管理成本约 20%（按单人工作量评估）。

(七) 某银行云成本优化实践

1. 企业需求和痛点

某股份制银行应用计划全部迁移上云，数据中心也在朝云化的数据中心转型。该客户依托现有的云平台，赋能开发和业务部门，实现自动运维，极大提升 IT 服务的灵活性，促进业务创新。

目前该客户本地私有云架构在 Azure 和 Hyber-V 上，公有云采用了 Azure 及 AWS 等主流公有云。容器平台架构在 Red OpenShift。随着该银行应用上云的逐步推进，给云资源管理带来了挑战，主要表现以下几个方面：

1) 支撑应用的云资源成本上升非常快：微服务和敏捷开发导致部署单元的急剧增长，缺乏针对云原生应用的资源管理方法和手段，无法做到对数以十万计的虚拟机和将来数以百万计的容器做全局的资源优化。

2) 存在应用资源容量过度配置的情况，没有达到按需配置、弹性伸缩的云建设目标。

3) 缺乏将应用与基础设施资源关联分析的手段，无法做到最优化使用资源，在保障业务系统性能的同时降低资源成本

2. 成本优化措施

经过探讨和验证，IBM 与客户针对云资源优化设计了整体方案，采用了 IBM Turbonomic 云资源管理平台结合以下应用场景进行云资源优化：

1) 自动可视化从应用到基础设施的依赖关系和资源情况

该客户应用繁多(近二百个)，依赖的资源错综复杂(近数十种)。对资源缺乏可视化管理手段。采用 Turbonomic 资源管理平台从应用视角出发，实时查看支撑的应用的运行状况，自上而下形成资源供应链，覆盖应用-服务-云-容器-虚拟化-主机-存储/网络完整资源链路。

2) 根据应用性能要求，自动发现资源瓶颈，自动生成云资源调整建议，改进应用性能。客户应用重度依赖云原生容器平台，容器平台资源

瓶颈直接影响应用性能。通过人工在海量动态的容器资源定位瓶颈和提出资源优化，可行性非常低。**Turbonomic** 资源管理平台结合应用性能监控平台，从应用性能驱动出发，快速定位影响应用性能的云资源瓶颈，并给出优化建议(比如扩展虚机的 CPU 或内存，或 POD 的资源)，从而提升应用性能。

3) 自动建议在不影响应用性能的前提下，如何优化利用资源，降低资源投入成本。客户反馈应用资源供应紧张,部门之间争用私有云 **Hyber-V** 资源冲突严重。但资源管理部门无法明确和衡量资源优化对应用的性能影响，导致资源优化管理落实困难。最后通过 **IBM** 云资源管理优化方案，在保障应用性能前提下，明确当前消耗资源数据，再基于 **AI** 给出优化计划和优化后节省的数据明细（比如当前使用 120 个 vm, 480 core CPU, 2 T memory, **Turbonomic** 自动监测并计算合适准确的资源负载为 100 个 vm, 360 core CPU, 1T memeory），减少推诿，提高部门合作效率。

4) 从自动的资源可视化和规划开始，可以逐步过渡到资源的自动化调整，全面提升自动化资源管理能力。客户初期对云资源优化解决方案的操作建议抱有不肯定态度。所以验证前期只提供建议，不执行，由客户手动处理。经过验证优化建议的准确性及对应用性能无影响，逐步接受自动化操作，实现全局自动化云资源优化。

3. 成本优化效果

通过 **IBM** 云资源管理解决方案帮助该客户实现：

1) 全面覆盖私有云容器资源、私有云虚拟化资源、公有云 IAAS 和 PAAS 资源，实现全局联动云成本优化。在保障应用性能前提下，节约优化成本近 25%。

2) 帮助快速解决故障，减少近 60%MTTR，节省了运维和 IT 支持与服务的成本，提升架构在云上的应用系统稳定和可用性。

通过 AI 赋能，提供可信的云资源优化操作建议和预警，确保优化操作合规和由被动响应转为主动预防。

(八) 某科研单位云成本优化实践

1. 企业需要与痛点

国内某科研单位电子档案库存储大量高价值数据，数据存量 300TB，每年数据增量 20TB，5 年内预计产生 400TB 数据，数据量较大且数据价值高需长期归档保存，数据一旦丢失将给业务系统带来巨大损失。这些归档数据大部分保存在磁带库中，磁带介质保存条件严格苛刻，而且磁带每 5 年左右需要进行数据迁移，在维持机房恒温恒湿环境和人工运维方面，原有的存储方式产生了大量的资金、人力资源浪费。该科研单位每年花费大量的资金存储归档数据，而这些归档数据访问频次极低，亟需寻求更加经济的云存储方案降低数据存储成本。新的存储方案要求：

1) 对 5 年内的存量数据和增量数据进行长期存储；

2) 新的存储方案需与原存储设备兼容，保证数据的安全性、一致性。

2. 成本优化措施

综合考虑成本、安全等因素，客户决定采用私有云蓝光存储进行冷数据归档存储。蓝光存储是以蓝光光盘为存储介质的一种新兴存储方式，具有存储周期长、低成本、高安全的特性，非常契合冷数据的长期存储场景。

蓝光存储主要从以下 4 个方面帮助客户实现云存储成本优化：

数据冷热分级存储：在客户本地部署蓝光网关，配置数据传输方式及生命周期策略，经常访问的热数据存储在与响应较快的磁性介质，不经常访问、需长期安全归档的数据存储在成本更低、更安全的蓝光存储，降低整体存储成本。

利用闲时网络：利用蓝光网关配置数据自动传输时间为晚间闲时，一方面不占用业务系统忙时带宽资源，另一方面利用闲时网络优惠达到降低成本的目的。

资源管理可视化、灵活扩容：首期为客户开通 400TB 存储空间，蓝光存储提供资源管理可视化控制台，客户随时掌握资源使用情况，当所需资源增加时，可根据客户需求灵活扩容；

兼容原存储设备：在客户本地配置蓝光网关，打通原存储系统与蓝光存储，无需增加系统改造投入。

3. 成本优化效果

1) 数据查询效率提升：优化前，客户随机读取磁带中数据，需倒带，查询效率低下；优化后，系统自动查找待查询数据所在光盘，查询速度快(平均查询耗时 1 分钟左右)；

2) 运维成本降低：优化前，离线存储速度远低于数据日增量，需大量近线存储阵列；需人工操作，整体维护成本投入大；优化后，蓝光存储能耗为原存储方式的 10%，大幅降低电费投入，存储速度有足够的余量满足数据日增量，极大节省近线存储空间；刻录和查询过程自动完成，降低人工成本。

3) 存储成本降低：根据数据的存储阶段配置生命周期策略，数据热温冷分级存储，相较不分级存储成本大大降低。

4) 资源管理更便捷：利用可视化管理控制台，随时掌握存储时长、容量、流量等数据，直观展示资源使用情况。

(九) 某保险企业云成本优化实践

1. 企业需求和痛点

某知名保险企业在推进数字化战略和一体化战略落地过程中，以 IT 变革为先导，深入应用云计算、大数据、人工智能等新技术，规划了新一代 IT 架构，全力推进 IT 架构从传统集中式架构向云化分布式架构转型。目前企业云已有 30000 多个虚拟机，包括腾讯云、华为云、青云、PowerVM、VMware 等公有云、专有云和虚拟化环境，支撑了各种保险业务系统稳定安全运行。其采用业界领先的云服务架构体系，构建了基于多中心多活的云服务模式，为新架构应用提供高可用高可靠的弹性基

础资源支撑。但是企业每年云成本投入面临如下痛点：

- 1) 每年采购大量设备，还是无法满足业务应用的需求，无法准确掌握业务资源的使用情况，提供主动性的保障计划；
- 2) 设备较多且分散在多个机房内，无法获取各机房资源的使用情况；
- 3) 无法预计设备的年增长率；
- 4) 无法统计各业务系统资源的利用情况，造成有的业务申请了大量的资源基本上未使用，而有的业务资源不足，造成了资源的浪费和资源分配不均；
- 5) 数据统计分析能力弱，无法及时提供领导决策用的统计分析数据。

2. 成本优化措施

企业采用服务商云管平台对企业的腾讯云、VMware、PowerVM 等云环境实现了纳管和监控，按业务和资源池维度对资源用量、资源性能、使用效率进行了充分的采集和分析，采用多种人工智能技术如：深度学习、ARIMA 算法、多目标求解和统计学模型等，为用户的成本优化提供了多种切实可行的建议方案，为保险企业整体的降本增效的发挥了巨大作用。主要通过以下手段实现整体云资源的成本优化。

1) 建立组织、项目组的配额体系。规划企业各部门和业务的资源配额，规范业务合理的资源用量，从预算和规划层面杜绝资源浪费。

2) 建立多维度、可视化的成本管理功能模块。通过自定义仪表盘体现组织和业务维度的资源配额、资源用量和资源效率的统计对比分析，提供具体的成本优化建议。直观呈现成本优化的关注重点和具体的调整

建议。

3) 提供整体规划层面的成本优化建议。提供新年度的 IT 资产采购建议方案。建议方案中在上年度云资源的增长率基础上，分析用户业务环境的资源规格比和物理资源规格比的差异（用户业务云主机的 CPU 和内存的配置比例为 1:8，而物理主机的 CPU 和内存的配置比例为 1:4，使得物理资源无法被业务应用充分利用），并结合企业的业务治理规范，通过多目标求解的人工智能算法模型，提供出针对性的资产采购建议和具体的资源池分配方案。在满足新年度用户业务增长的前提下，弥合用户环境中的物理资源规格比和业务资源规格比的差别，提升物理资产成本的使用效能，实现在整体规划层面的降本增效的效果。

4) 利用深度学习算法和 ARIMA 统计模型的云资源增长和业务变化趋势预测。自定义划分维度、设定管理红线、根据云环境/业务资源的历史年增长率，结合资源分配原则和使用情况计算出未来 1-2 年的容量需求。

5) 高效的数据统计分析。数据结构优化，通过定时对数据进行清洗聚合，提供标准化的明细、按小时、按天聚合的数据，显著提高数据统计分析效率。

3. 成本优化效果

1) 根据组织和业务的实际资源用量，调整组织、项目组的配额，在保证业务应用的所需要的云资源的前提下，通过组织业务的配额调整手段，结余出 23% 的云资源支撑新业务的开展，显著增强了云平台的业务

支撑能力。

2) 通过闲置云主机的调整和配置优化，提高了资源使用效率 45%，节约了 32% 的成本支出。

3) 通过对云环境的年度资源用量增长情况的历史数据统计分析，可以提供合理的新年度的资产采购建议，在此过程中弥合物理资源规格比和业务资源规格比的差别。可显著地优化云成本的投入和提升了 25% 的资源效益。

(十) 某广告公司广告平台云成本优化实践

1. 企业需求和痛点

Mintegral 是汇量科技 Mobvista 旗下移动广告平台，致力于运用其流量优势及广告资源，帮助全球各个地区的开发者与发布商提升用户规模与变现收益。Mintegral 每天的广告峰值请求量达 1000 亿次以上，同时，平台需对单次广告请求的实时性要求很高，需做出毫秒级的精准预估和出价回应。基于全球化的业务定位，Mintegral 的系统全部建立在公有云上（AWS、阿里云）。

移动广告业务具有很强的潮汐特色，Mintegral 的峰值流量是均值流量的 4 倍。这一场景下，虽然可以根据流量变化在云端弹性伸缩计算资源配合按需实例，实现计算资源和业务量的动态匹配，但按需实例的高昂价格，使得在实际情况中，Mintegral 不得不使用相当数量的预留实例；而预留实例的静态特性，又不可避免地会带来资源和成本的浪费。

Mintegral 的云资源配置方案为：25%按需实例，75%预留实例。（按需实例价格最贵，可随时拉取、即买即用；而预留实例价格为按需实例的 5-6 折，但需承诺 1-3 年的月使用量，相对静态，难以适应业务量的波动。）为帮助 Mintegral 解决用云成本问题，Mobvista 技术平台自研了一套云资源优化解决方案，后将其产品化为 SpotMax。

2. 成本优化措施

在云上要实现成本优化，最根本的就是要充分利用云计算的特点，尤其是：“按需获取”，“按使用付费”。

1) 由单体架构到微服务架构

Mintegral 原有的应用程序由大型单体系统组成，所以即使是系统中仅某个功能达到瓶颈，由于是单体系统，在横向伸缩时也需要复制整个单体系统，因此需要较大机型来运行。为解决这一问题，技术团队将其重构为微服务方法，可以精准地扩展瓶颈所在微服务，以便使用更小更适合的机型精确地实现系统的伸缩。此外，云端具有“按使用付费”的特性，大机型和小机型存在很大的价格差，精准伸缩就带来了计算成本的节省。

2) 利用竞价实例

如上所述，Mintegral 业务的具有潮汐特色，按需实例的灵活性虽然可以有效适配这种场景，但其价格太贵。因此，技术团队选择了云上的另一种鲜少被大家所广泛使用的收费模型--竞价实例。竞价实例有着和

按需实例一样的灵活性，而其价格通常仅为按需实例的 30%，最低可达按需实例的 10%。

同时，使用竞价实例，也需面对以下挑战：**a.** 根据不同类型，不同区域实例供需关系的变化，正在使用的竞价实例有可能会被回收，影响服务的稳定性。**b.** 根据供需关系变化，紧缺的实例类型可能无法获取，影响系统的伸缩。

为了克服上述问题，以帮助 **Mintegral** 不仅在离线的大数据部分，同时也能够在线上处理系统中广泛使用竞价实例，团队从如下 5 个方面入手，构建了自动化系统来进行竞价实例资源的管理和调度：

1) 在集群中合理配置多种实例，防止竞价实例回收导致集群出现大比例的计算资源损失。该系统通过跟踪收集不同类型、不同区域竞价实例的回收情况，并经过大数据分析，结合历史及实时的资源情况，实现集群中竞价实例类型的合理配置。

2) 利用大数据及 **AI** 技术，持续优化集群，降低集群中断概率。利用大数据及 **AI** 技术，对集群中不同类型实例的中断率进行实时预估，并以低中断率类型实例替换高中断率类型实例，从而整体降低集群中实例中断的概率。

3) 在资源不足时，自动选择兼容类型或按需实例进行集群扩展。集群按需扩展时，若面临资源紧缺、无法获取设定的竞价实例，我们构建的系统将自动获取兼容类型的竞价实例或者按需实例，以保证集群顺利进行扩展。

4) 积极的中断预处理。使用当中的竞价实例即将被回收前，系统将接到云商预警通知，并提前向集群补充实例，以保持实例被回收后集群的计算容量，从而保证服务质量。

5) 按需实例打底。系统在无法获取竞价实例及可使用的兼容实例时，会自动获取相应的按需实例来补充计算资源，并且在需求关系变化后、所需的竞价计算资源可用时，系统会自动将按需实例移除，替换回竞价实例资源，以实现成本的节省。

3. 成本优化效果

经上述方案进行优化后，与 2017 年 12 月相比，Mintegral 在承载不断增长的广告请求数量的同时，其单次广告请求的成本降低了 65%。此外，为了能够让其他企业也能更好地利用竞价实例、降低用云成本，技术团队提升了该云成本优化内部系统的适用性并将其产品化，形成了 SpotMax 这套基于云原生的解决方案。

以 2021 年全年的优化成果为例，SpotMax 成功为 Mintegral 解决实例回收问题 72438 次、处理资源获取失败问题 3972 次。在 Mintegral 业务大比例使用 Spot 实例的情况下，从未发生因 Spot 实例回收导致的严重线上事故。



中国信息通信研究院

地 址：北京市海淀区花园北路52号

邮政编码：100191

联系电话：010-62300559

传 真：010-62304980