

构建高性价比的私有云

2022年3月

简介

通用私有云基础架构的成功主要取决于其性价比。对于公有云基础架构而言，拥有良好架构的私有云可以是一种具有成本效益的扩展架构，可以确保最大限度地减少多云成本。多云架构可以帮助企业将其工作负载始终托管于最具成本效益的地方。这反过来又降低了长期总体拥有成本（TCO），为创新和增长留出更多预算。

同时，在构建私有云时，决策出错可能会对业务产生负面影响。如果私有云的架构较差，可能导致其性能下降，使其不适合其目的。反过来，如果为了优化性能而添加更昂贵的组件，可能很快就会超出预算。因此，我们可能会发现正在构建的私有云根本不具备成本效益，并且可能无法满足企业的理想投资回报率（ROI）目标。

这就是为什么我们要构建一个高性价比的私有云。尽管为了实现私有云，企业需要预先付出大量的资本支出，但仍然可以在确保所需容量的情况下，以最高的性价比对其进行配置。最佳的私有云架构应进一步缩减许可和订阅成本，以及实现基础架构部署和后期部署运行的完全自动化，这样才可以达到最大的资本支出和运营支出效率。只要控制住了预算，企业就能以更优惠的价格提供服务，或将盈余预算投资于研究创新。

本白皮书展示了Canonical构建高性价比私有云的最佳实践。我们认为，企业应谨慎挑选硬件，实现高性能的基础架构，同时确保将资本支出和运营支出的开销降至最低。随后，我们提供了一个Canonical内部云的案例研究，包括根据这些最佳实践部署的性能结果。最后，我们根据该云的硬件物料清单（BOM）进行了性价比分析，并给出了每台虚拟机（VM）的成本。

在构建私有云时确保高性价比

构建一个高性价比的私有云首先从硬件层面开始。尽管其他成本（许可、订阅、运行等）也必须考虑在内，但这些通常取决于供应商和客户，而云架构是通用的。因此，对于我们的参考架构，我们主要关注硬件选择和配置指南，以最大限度地提高私有云的性能，同时将硬件成本降到最低。

概述

如图1所示，一般基础架构预算包括四大支柱：硬件设施、软件许可和服务、内部运行与维护以及研究和开发。前三项属于原始成本，而后一项可帮助企业探索和采用新技术，促进企业转型，从而增强其市场竞争力。因此，每个企业都在不断增加其研发和业务转型成本投入。

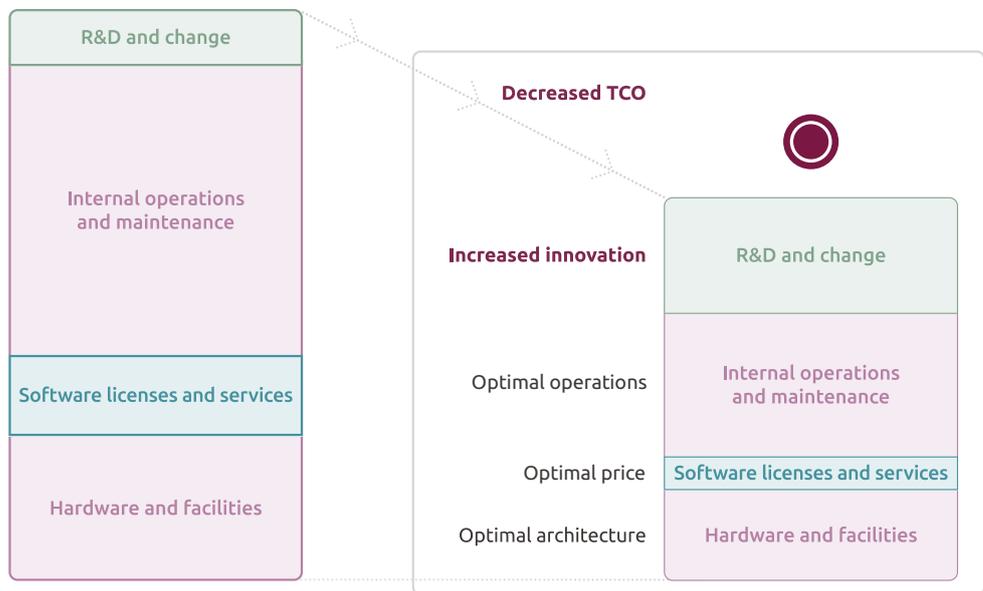


图1，典型的基础架构预算。

另一方面，预算控制也很关键。硬件、软件和运行成本都给内部信息技术（IT）团队带来了许多压力。因此，为了增加研发与变革上的预算，他们不得不减少其他支出。但是，最好不要降低平台质量，因为基础架构质量不佳会直接影响在其上运行的工作负载的质量，从而导致企业提供的服务质量差、客户不满意、索赔等。

因此，为了降低基础架构的TCO，同时增加研发与变革的预算，企业必须为最佳架构、最佳运行和最优服务成本设计基础架构。最佳即尽可能便宜，但又不影响质量。获得最优服务成本非常简单，只需要选择对供应商，即选择采用市场上最经济透明的价格结构的供应商。

要实现最佳运行，企业需要选择一个在部署后无需过多交互的平台（即提供全自动化运行）。反过来，要实现最佳架构，需要在选择硬件和软件组件时应用适当的方法，并将它们配置好。

方法

选择硬件和配置时，通常从选择计算、网络和存储资源开始。难点在于设计一个可以实现所需容量和性能，并最大程度地减少基础架构预算压力的架构。通过分析工作负载的需求，以及为所需的中央处理器（CPU）线程、随机存取存储器（RAM）和存储量设计云，可以相对轻松地解决容量需求；不过，处理性能需求并非易事。尽管高端硬件通常能够达到所需性能，但是可能会导致预算不足。而低端硬件可能无法达到企业的性能要求。

但是云的性能与价格之间并不是线性的关系。根据Canonical的经验和观察，它们之间的关系很像一条对数曲线，如图2a所示。这是因为相比于用于管理虚拟化堆栈并提供软件定义网络（SDN）和软件定义存储（SDS）抽象的云软件，当前硬件的能力远远超过这些云软件的能力。因此从某个角度而言，使用更强大的硬件并没有意义，因为它们不会提高云的性能，反而会给基础架构预算带去不必要的压力。这就是为什么我们需要找到一个满足两者需求的最佳架构。为了促进这一过程，Canonical不断地对各种云架构进行基准测试，并进行性价比分析，以找到最佳的架构。

一旦确定了最佳架构，下一个挑战便是确定用于构建云的组件的最佳大小。组件大小通常不会对云的性能产生影响，不过可能会对云的价格产生影响。以存储为例，存储设备的价格因其大小而异，但价格与大小之间的关系并不是线性的。两者之间的关系更像是一条U形曲线，如图2b所示。因此，企业需要找到何时使用较大磁盘会导致我们丧失成本效益，这一点很重要。另一方面，使用较小磁盘需要更多设备来满足云容量需求。如此一来，成本上升，往往还需要更多服务器，进而导致总硬件成本增加。

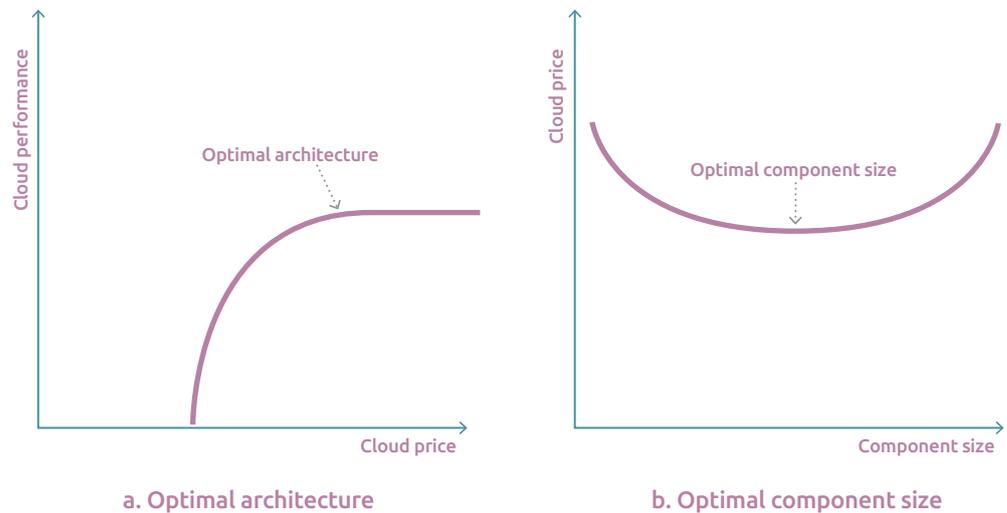


图2，私有云成本分析。

最佳组件大小一般处于曲线的中间位置。这一点同样适用于其他主要硬件组件，即RAM。组件的价格在不断变化，因此，每次构建新的云时都应该执行这个过程。Canonical定期执行这一过程，并不断更新其私有云实现的参考架构，以确保在可用硬件的基础上实现最高性价比。

其他考量

除了遵循基于U型曲线的方法之外，在构建有性价比的私有云时，还必须考虑以下因素。

云架构

云架构的选择通常是企业在设计云时需要做出的第一个决定。从性价比的角度来看，这也是一个重要的决定。架构决定了将部署多少节点，以及云服务将如何在这些节点之间分布。虽然为单个服务运行单独的节点通常会实现更佳的性能，但会对基础架构的预算造成不利影响。在构建高性价比私有云方面，Canonical拥有经过验证的方法，即使用超融合架构。该架构假设所有节点都是平等的，并在整个集群中均匀分布计算、网络、存储和控制平面服务。这种方法确保了同质性和更高的资源利用率，且与其他私有云实现架构相比，硬件成本更低。

服务器类型

一旦决定了选择哪种CPU、RAM和磁盘，下一步就是选择一个服务器作为资源的底盘。市场上有各种类型的服务器，需要考虑许多不同的参数。然而，从性价比的角度来看，最重要的参数是规格。如果用的服务器太大，它们占用的空间就更多。也意味着要用到更多机架，增加了总硬件成本。而如果服务器太小，可能就没有足够的空间来容纳足够数量的PCIe插槽。

因此，Canonical建议使用2U服务器来实现私有云。2U服务器确保了总硬件成本和云性能之间的平衡。这类服务器能容纳更高端的CPU、额外的网卡（NIC）、非易失性内存主机控制器接口规范（NVMe）设备和其他的卡，同时不会给基础架构预算带来巨大的压力。

存储缓存

在为私有云设计存储时，仅仅选择最佳磁盘大小是不够的。这是因为有各种类型的磁盘：硬盘驱动器（HDD）、固态硬盘（SSD、英特尔傲腾和各种类型的接口：序列式SCSI（SAS）和NVMe。它们每TB的成本相差了几个数量级。传统的SAS HDD价格便宜，但性能有限。而如SSD和英特尔傲腾这样的NVMe设备拥有更出色的性能，但价格要贵得多。这种情况下，实现分层存储是最折中的解决方案。虽然HDD磁盘以低成本提供了所需的存储容量，但NVMe设备用于在实际存储之前实现缓存层。该缓存层不会大幅增加总硬件成本，同时显著提高了云性能。

网络拓扑

通过存储缓存技术解决了机器层面的所有潜在瓶颈之后，还要解决网络层面的瓶颈，这也很重要。Canonical私有云实现的参考架构使用链路聚合控制协议

（LACP）捆绑链路，其最低速度为10 Gbps，以确保云中所有节点之间的高吞吐量。此外，它基于两层叶脊网络拓扑结构，通过最大限度减少跃点数来降低网络连接的延迟。通过尽可能地进行基于以太网帧的数据包转发，叶脊拓扑最大限度地提高了吞吐量，免去了网际协议（IP）头域检查的支出。这样就建立一个超快的网络结构，能够包含云的所有节点，更重要的是，还包括了所有的虚拟机。尽管快速网络会增加总硬件成本，但在对性能敏感的云环境中，这样的扩展是非常必要的。与其他硬件组件（即服务器）的成本相比，这些成本并不算高。

性能扩展

最后，如果需要的话，还应该考虑额外的性能扩展。Canonical支持所有常见的性能扩展，包括单根输入/输出虚拟化（SR-IOV）、数据平面开发套件（DPDK）、CPU亲和性、非均匀访存模型（NUMA）、大内存页和PCI直通等，这些通常都是在性能敏感环境中需要的。由于这些扩展可能会给基础架构预算带来额外的压力，因此建议在设计阶段将其作为需求提出，以便Canonical能够帮助选择硬件。

私有云性价比分析

设计私有云实现的最佳架构需要仔细选择硬件，并在性能和价格之间找到平衡，同时确保云的预期容量。然而，企业还应该为实现最佳运行和最优价格（图1所示）设计他们的基础架构，以降低与云基础架构维护相关的总体TCO。

确保最高的资本支出和运营支出效率

选择最佳架构可确保资本支出成本的最高效率，而为了最佳运行和最优价格设计云有助于降低每年的运营支出成本。尽管选择硬件很重要，其价格通常是初始投资的重要组成部分，但企业不应低估与云的后期部署维护相关的经常性成本。其中包括软件许可证和服务以及内部运行。如果选择了购买和续期许可证都很昂贵的平台，或者需要整个团队运行，TCO将很快飞快增长。因此，在做出选择时，企业应该考虑到代码开放性、定价结构透明度和运行自动化等问题。优化上述因素可以帮助预测预算，并有助于避免后期部署成本高昂的意外情况。

开放式基础架构堆栈

当谈及私有云实现时，Canonical利用了所谓的开放式基础架构堆栈。堆栈（图3所示）由开源技术组成，覆盖了基础架构中从裸机到微服务的所有层。Canonical的方法是利用最好的技术，让其发挥所长，并在单一的订阅中提供全栈支持。开放式基础架构堆栈的核心是OpenStack。OpenStack是实现私有云最流行的开源平台，OpenStack管理分布式计算、网络和存储资源，并借助虚拟化技术，通过自助服务门户将这些资源分配给虚拟机。OpenStack可以轻松通过基于Kubernetes的容器化层进行扩展，并在其上运行。

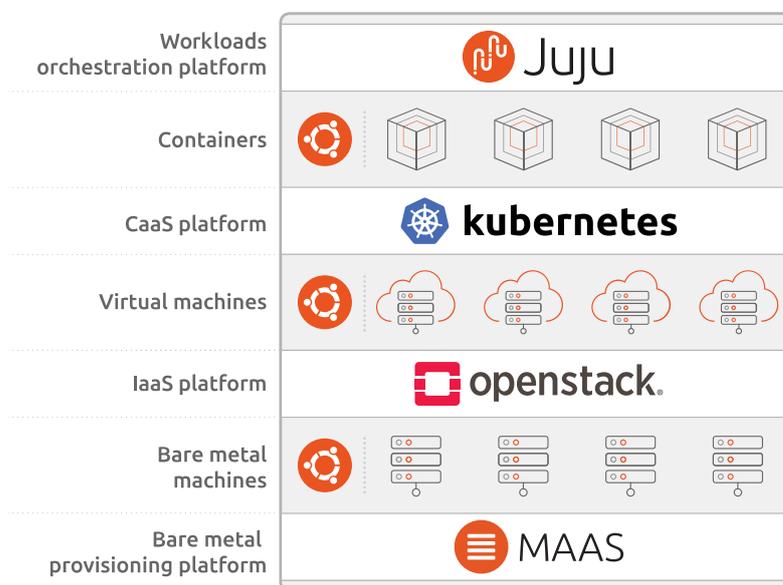


图3，开放式基础架构堆栈。

针对Canonical开放式基础架构堆栈的四个层级的商业支持：

基础	包括安全更新、强化和内核Livepatch
标准	在正常办公时间通过电话和工单支持扩展基础层级服务
高级	通过全天候支持和更积极的SLA扩展标准层级服务
全托管	整个堆栈由Canonical的云专家团队监控、管理和运行，包括解决事件和问题、升级以及日常操作

针对私有云实现的Canonical服务

由于采用开源技术存在诸多挑战，Canonical为企业客户提供广泛的商业服务。其中包括咨询服务、商业支持和全面代管服务。使用上述服务可以确保企业不会陷入OpenStack和Kubernetes部署。同时也能在必须满足积极的可及性目标和服务水平协议（SLA）的生产中，放心运行整个开放式基础架构堆栈。采用Canonical商业服务的客户包括MTS、BNP Paribas、Best Buy和Bloomberg。

案例研究——Canonical的内部云

为分析Canonical用于私有云实现的参考架构性价比，基于呈现的指南设计并实现了一个示例云。

示例场景

PCB和PCB Plus是Canonical的固定价格咨询包，针对的是基于参考架构和认证硬件的私有云实施服务。

服务包括云设计和交付、on-prem工作坊、工作负载分析和迁移计划。

两个包均使用了Canonical的私有云实现工具集，包括裸机即服务（MAAS）、Charmed Operator Lifecycle Manager（OLM）和OpenStack Charms。

作为私有云构建（PCB）服务的一部分，使用Canonical的私有云实现内部工具部署了云。云已部署在一个数据中心，由37个云节点和另外3个用于云自动化和维护的基础架构节点组成。为确保资源利用最大化，其选择了超融合架构。表1列出了云和基础架构节点的硬件规格。软件规格如表2所示。

云节点 (37 x GB- R282-Z90)	
规格	2U 机架, LFF 驱动器笼
处理器	2 x AMD EPYC Rome 7502 (32c/64t 2.5GHz)
内存	32 x 64GB 3200MHz DDR4 RDIMM (共2048GB)
存储	2 x 6.4TB U.2 NVMe (三星 PM1725b) 6 x 12TB SAS 3.5" HDD 4 x 英特尔® 傲腾™ NVMe SSD 900P 480GB
网络	1 x 双口 100 GbE NIC (Mellanox ConnectX-5 Ex QSFP28) 1 x 双口 1 GbE NIC (英特尔® I350-AM) 1 x BMC
基础设施节点 (3 x HPE Gen10+ DL325)	
规格	1U 机架, SFF 驱动器笼
处理器	1x AMD EPYC 7302 (16c/32t 3.0GHz)
内存	2 x 64GB 3200MHz DDR4 RDIMM
存储	4 x 4TB SAS 2.5" SSD 硬RAID控制器
网络	1 x 4口 1 GbE NIC (英特尔® i350-T4) 1 x BMC

表1, Canonical内部云的硬件规格。

Canonical的内部云	
Operating System	Ubuntu 20.04 LTS
Kernel version	5.4
OpenStack version	Ussuri

表2, Canonical内部云的软件规格。

示例定价

由于同一资产的价格可能因供应商而异，Canonical进行了招标以选择性价比最高的硬件。在招标中，所需的云节点规格分布在不同的硬件供应商之中，最终在需求匹配的供应商中选择了报价最低的。Canonical以大约120万美元的价格为这个有40个节点的云购买了硬件。部署的云已经为74个霄龙ROME 7502 CPU提供服务，每个CPU有32个核心、74 TB RAM、229.4 TB临时存储和2664 TB原始持久存储，能容纳数千个资源丰富的虚拟机。

性能基准

为了衡量云的性能，通过灵活的输入/输出测试工具（fio）在云上运行了几个测试。在这些测试中，用了200个客户实例在云上进行随机读/写运行，以确保存储的压力足够大。每个实例使用了8个vCPU，以确保在物理CPU和2个持久存储设备之间保持平衡分布，每个存储设备的容量为200GB。为了确保在测试期间可以超过缓存设备的总容量，可寻址存储的总容量被缩小到80TB。每次测试持续1小时，在15分钟内客户实例从0上升到200并返回到0，以避免同时让所有存储设备超载。为了测量总的每秒输入/输出操作（IOPS）容量，使用了4KB随机数据工作负载。为了测量总吞吐量，使用了4MB顺序数据工作负载。所有测试都在新部署的云上进行，并重复多次以确保结果一致。

基准结果

表3为收集到的所有指标。根据数据显示，Canonical的内部云实现了与其他关于私有云性能的白皮书类似的性能结果[1]。由于这两个案例都使用了相同的底层存储平台，实验室条件也相似，因此得出的结论是，Canonical构建私有云的方法既可以满足性能需求，也可以控制预算。

4KB 随机工作负载		
指标	写入	读取
平均吞吐量 [KIOPS]	450	10240
平均第 99% 的延迟 [ms]	38	2
平均延迟 [ms]	30	1
4MB 顺序工作负载		
指标	写入	读取
平均吞吐量 [GB/s]	12	258

表3，性能结果来自Canonical的内部云。

TCO 分析

尽管在选择私有云实现的平台时需要考虑许多因素，但每个虚拟机的成本通常是决定性的因素。不同的平台基于不同的架构，在软件许可证和服务方面有着不同的定价结构，并且对平台的后期部署运行有不同的要求。因此，即便是相同大小的不同平台，与云基础架构维护相关的TCO也会不同。而它们容许的容量一样。因此，每个虚拟机的成本也因平台而异

vCPU 数量	内存容量 [GB]	临时存储容量[GB]	持久存储容量[GB]	理论最大虚拟数量 [VM]	每台虚拟机每小时成本 [美元]
1	4	8	40	11544	0.0092
2	8	8	80	5772	0.0184
4	16	8	160	2886	0.0369
8	32	8	320	1443	0.0738
16	64	8	640	721	0.1477

表4 Canonical内部云中每台虚拟机成本指标。

表4显示了来自Canonical内部云的每台虚拟机每小时成本指标。这些估计的数字包括了与私有云部署和运行相关的所有必要成本，包括硬件成本（机架、服务器、交换机、布线等）、托管服务（租金、电费、网络和硬件维护费）、云交付和年度订阅（假设由Canonical提供全面代管服务）。此外，在进行这些计算时，我们假设硬件更新周期为3年，CPU超量使用比为2:1，RAM无超量使用，默认的Ceph复制因子为3，云利用率为75%。

结语

尽管构建性价比是一项艰巨的任务，但遵循某些指南和最佳实践有助于最大程度地减少TCO，同时不影响基础架构的质量。由于总硬件成本很高，私有云实现需要大量的前期投资，因此设计一个最佳的架构是企业应该关注的第一步。在各种考虑因素中，Canonical在选择云资源和应用存储缓存技术时遵循一定的方法，这是其实现高性能的云的有效方法，并没有给基础架构预算带来压力。

在本白皮书中，我们呈现了Canonical的私有云实现参考架构如何满足性价比的需求。根据“TCO分析”一节中提供的数据，与领先的公共云和私有云提供商相比，Canonical内部云中每台虚拟机的成本明显要更低。我们还强调，选择一个基于开源技术、实现透明定价结构和全自动化运行的平台有助于降低经常性成本。

下一步

与您的Canonical销售代表联系，了解私有云实现的下一步步骤。Canonical维护一个参考架构和参考硬件文档，概述了高性价比私有云实现的详细建议。我们强烈建议您遵循这些建议，因为设计阶段的错误决策可能会对云的性能和基础架构预算产生不利影响。

如果您还没有与Canonical联系，可以通过<https://cn.ubuntu.com/contact> 联系我们。

1. <https://www.samsung.com/semiconductor/global.semi/file/resource/2020/05/redhat-ceph-whitepaper-0521.pdf>