

从微观经济学看云计算市场发展

文 / 蒋清野, Young Choon Lee, Albert Y. Zomaya

基于2006到2013年的历史市场数据, 本文从微观经济学的角度分析了企业计算资源市场的供需关系。工业界可以依据本文所提供的数据设计更好的定价策略, 从而同时达到市场扩张和增加营收的目的。

在一个充分竞争的市场环境中, 没有一个参与者强大到足以单方面控制某一类产品的价格。在这种情况下, 细微的价格调整也能对产品销售带来巨大影响, 并进一步影响产品的市场份额和营收情况。在微观经济学中, 用需求价格弹性来描述价格变化对产品需求的影响。用户对价格变化非常敏感的需求被认为是柔性需求, 反之被认为刚性需求。

在企业计算资源市场中, 同时存在多个竞争对手, 通常被认为是一个充分竞争的市场环境。在前些年, 企业或者直接从硬件厂商购买服务器, 或者从托管服务提供商那里租赁服务器。近几年来, 公有云服务逐步进入企业计算资源市场, 其市场份额呈稳步上升的趋势。在美国以及世界上许多地区, 亚马逊网络服务(AWS)被广泛认为是公有云服务的代表。在过去几年当中, 市场上涌现了大量基于AWS开发的新应用, 同时有大量已有应用迁移到AWS上。以直接向互联网提供Web服务的服务器数量来计算, AWS已成为世界上最大的服务器托管公司。由于AWS所取得的巨大成功, 市场上很快出现了一些追随者, 包括Windows Azure、GCE、HP、Rackspace等。这些新的公有云服务提供商进入市场后, 通常采用低价作为吸引用户的手段, 而现有的服务提供商则被迫通过降价作为回应。价格战争引发了一系列与公有云服务相关的讨论, 甚至延伸到公有云服务会不会逐步免费。

企业计算资源市场和需求价格弹性

企业计算资源市场

目前, 企业获取计算资源的主要途径有三种。它们可以从硬件厂商购买新的服务器, 从托管服务提供商那里租赁服务器, 或者是使用公有云服务。这三个途径代表了企业计算资源市场的三类业务, 也就是服务器销售业务、服务器租赁业务和公有云服务。对于企业来说, 不同的业务有不同的供应商, 不同的财务、时间和人力要求, 以及不同的资源生命周期。

服务器销售业务的主要供应商是服务器厂商, 包括DELL、Fujitsu、HP、IBM和Oracle。用户既可以直接向服务器厂商下单, 也可以通过第三方渠道或者代理商下单。购买服务器的费用在财务上被计入固定资产开支(Capital Expense, CapEx), 通常是一次性支付给供应商的。标配的低端服务器通常都有本地库存, 资源交付时间可以控制在一两周以内。高端服务器往往需要提前下单, 交付时间有可能达到数月之久。服务器被交付给用户时, 其所有权也同时被转移给用户, 它们便成了企业的用户资产。服务器硬件的生命周期通常为三到五年, 具体的时间则取决于用户的固定资产折旧计算方法。

服务器租赁业务的主要供应商是托管服务提供商, 包括Rackspace、OVH和SoftLayer。用户既能直接和服务提供商签订服务合同, 也能通过代理商签订合同。根据企业对计算资源的需求, 合同

的跨度从几个月到几年不等。依据服务合同，用户定期向服务提供商支付服务费用以维持通过网络使用特定计算资源的权利。租赁服务器费用在财务上被计入运营开支（Operations Expense, OpEx）。资源交付时间通常在数天以内，但也有可能由于服务提供商需要订购新硬件的原因延长到数周之久。通常，服务器租赁合同并不改变计算资源的所有权。有些服务提供商会选择一个长期合同（如三年）执行完毕后将计算资源的所有权赠送给用户。

公有云服务主要供应商是AWS、GCE、Rackspace和Windows Azure。用户能通过门户网站或API直接请求并获得计算资源。类似于水电等公共资源服务，用户只需为它们所使用的计算资源（包括数量和时间）付费，相关费用在财务上被计入运营开支。资源交付时间通常在几分钟左右，而技术进步正在将这个时间缩短到秒级。对公有云服务来说，计算资源的所有权是属于服务提供商的。除按使用量计费外，用户无须为底层基础设施承担任何费用。

需求价格弹性

1890年，现代经济学先驱Alfred Marshall发表了他的著作《Principles of Economics》。按照Alfred在该书中提出的概念，需求价格弹性等于需求变化百分数与价格变化百分数的比值，即：

$$E = \frac{(dQ/Q)}{(dP/P)}$$

其中，E代表需求价格弹性，P代表价格，Q代表销售数量，dP和dQ分别代表价格和销售数量变化的百分数（取绝对值）。当E大于1时，需求变化百分数大于价格变化百分数，表示需求为柔性需求。当E小于1时，需求变化百分数小于价格变化百分数，表示需求为刚性需求。当E等于1时，需求变化百分数等于价格变化百分数，表示需求为单一需求。显而易见，需求价格弹性可以量化衡量用户对价格的敏感程度。此外，需求价格弹性还能直接影响营收。对于刚性需求，价格上涨导致营收增加。对于柔性需求，价格下降导致营收增加。

产品的需求价格弹性不是一个常量，而是一个与价格、需求及考察时间密切相关的变量。一般

来说，有两个因素对产品的需求价格弹性影响重大：产品的可替代性，如果用户有多个可替代产品，则需求趋于柔性；用户用于该产品的预算占用户总预算的比例，这个比例越大，用户对价格变化越为敏感，则需求趋于柔性。

数据和讨论

数据来源

在此项研究中，采用了2006到2013年的历史市场数据作为分析依据。对于服务器销售业务，采用Gartner关于全球服务器出货情况季度报告中的数据。对于服务器租赁业务，以Rackspace作为代表，并从Rackspace向美国证券交易委员会（SEC）提交的10-K年报中提取数据。对于公有云服务，选择AWS EC2为代表，并从AWS的官方博客网站获得AWS EC2的历史价格数据。AWS的规模一直是亚马逊的顶级机密之一。在此项研究中，我们综合了一些业界知名人士对AWS规模的估算数据，并使用这些估算数据作为讨论依据。

市场整体状况

图1为2006到2013年间全球服务器出货数量、Rackspace公司的服务器数量和AWS EC2服务的物理服务器估算数值。显而易见，服务器销售业务的需求远远大于服务器租赁业务和公有云服务的需求。但服务器销售业务和服务器租赁业务的增长近似于线性增长，而公有云服务则呈现指数增长的趋势。换句话说，公有云服务的需求增长速度远远大于购买服务器和租赁服务器的需求增长速度。

服务器销售业务

图2为2006到2013年间全球服务器出货数量和营收

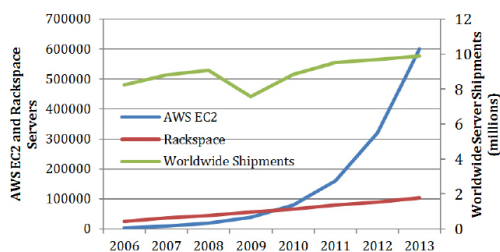


图1 全球服务器出货数量、Rackspace服务器数量、AWS EC2服务器数量增长

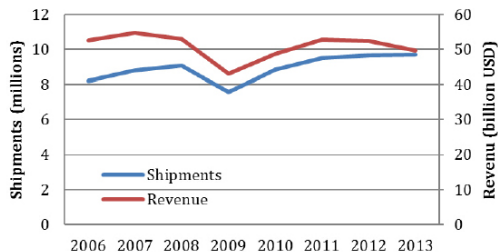


图2 全球服务器出货数量和营收数额

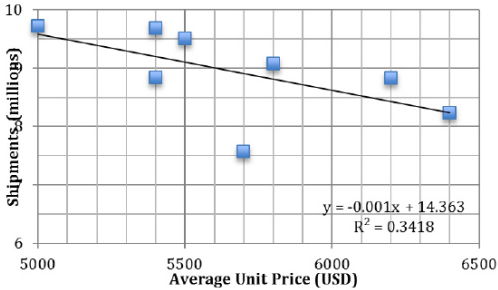


图3 全球服务器出货数量和平均价格

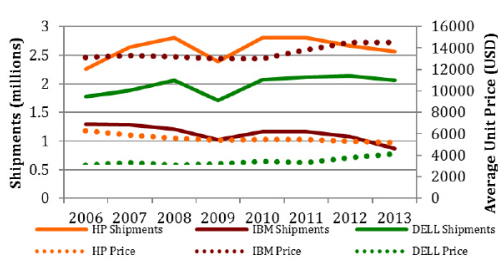


图4 HP、IBM和DELL的出货数量和平均价格

数额。2009年出货数量和营收数额的下跌，显然是2008到2009年全球经济危机的结果。尽管服务器硬件在配置上（如处理器、内存、存储）各有差异，而我们假设每台服务器在统计意义上具有相同的属性和平均价格，可以被作为相同的产品看待。图3为服务器销售业务的供需数据，横轴代表平均价格，纵轴代表销售数量。根据线性拟合结果，可以算出2013年全球服务器销售业务的需求价格弹性为0.51，属于很强的刚性需求。

对大部分企业用户来说，当需要更多计算资源时，首先想到的是购买一些新的服务器。其他形式的计算资源由于种种原因不能完全替代购买服务器：个人计算机处理能力非常有限，超级计算机价格过于昂贵，租赁服务器可能会存在安全问题，公有云服务尚未得到普及。此外，购买服务器硬件属于固定资产投资，其预算通常是在整个公司层面决定的，并且只占公司年度预算的一小部分。缺乏可替代品和预算比例很小共同导致了需求的刚性。也就是说，企业根据其业务规划决定服务器

硬件投资，基本上不受市场上价格波动的影响。

刚性需求的一个根本特征是需求变化的百分比小于价格变化的百分比。因此，下调价格在带来销量的同时会导致营收下降。在一个垄断行业中，供应商往往会选择反其道而行之——提高价格，降低销量，增加营收。问题在于，服务器硬件行业的竞争非常激烈，即使是领先的厂商也通过下调价格来换取市场份额。如图2所示，尽管全球服务器的出货数量在增加，但总营收数额看起来反而有下降的趋势。这意味这服务器业务的利润越来越薄，服务器厂商的日子越来越难过。

图4为三家主要的服务器硬件厂商（IBM、HP、DELL）在2006到2013年间的出货数量（实线）和营收数额（虚线）。不同厂商的平均价格之间存在巨大差异，意味着这些厂商并不是在同一个细分市场里相互竞争。IBM服务器的平均价格大于13000美元，属于高端服务器；HP服务器的平均价格在6000美元左右，属于中端服务器；DELL服务器的平均价格低于4000美元，属于低端服务器。IBM服务器的平均价格在2006到2010年间基本上没有变化，但其出货数量一直在下降。可以认为，需求的减少并不是市场对价格调整的反应，而是市场的整体偏好正在从高端服务器往低端服务器（很有可能是从Power服务器往x86_64服务器）迁移（2008到2009年全球经济危机对出货数量的下降也有影响）。2010到2012年间，IBM服务器的平均价格上调了11%，出货数量下降了8%。由此算出在此期间IBM服务器的需求价格弹性为0.73，属于刚性需求（我们有意忽略了2013年的数据，因为这一年IBM正在设法出售其服务器产品线，对相关业务有很大的影响）。2006到2013年间，HP服务器的平均价格下降了18%，出货数量增加了13%。由此算出在此期间HP服务器的需求价格弹性为0.72，与IBM服务器非常类似，属于刚性需求。在同一时间范围内，DELL服务器的平均价格上升了28%，出货数量也增加了15%。平均价格和出货数量同时上升，意味着需求的刚性很强。

需求的刚性及平均价格之间的巨大差异，表明IBM、HP、DELL的服务器并不是能互相替代的。从经济学角度来看，这些厂商在各自细分市场里形成了事实上的垄断，具备控制定价的能力。因此，市

场偏好及其自身定价策略对营收数额的影响要远大于来自竞争对手的影响。图5为IBM、HP、DELL的服务器业务营收数额。IBM的提价策略是对刚性需求的正确反应。但市场的偏好正在从高端服务器往低端服务器迁移，导致服务器业务营收的持续下降。市场偏好的迁移，也许是2014年初IBM将其服务器业务出售给联想的根本原因。HP的降价策略是对刚性需求的错误反应。降价在提高出货数量的同时导致营收数额降低，并导致利润率的降低。为了应对这样的危机，HP于2012年4月发布HP Cloud试水公有云服务。DELL的提价策略是对刚性需求的正确反应。作为市场偏好迁移的受益者，DELL同时实现了出货数量和营收数额的增长。

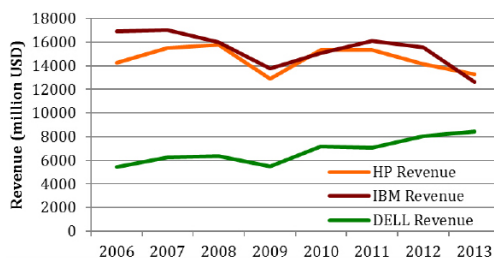


图5 HP、IBM和DELL的服务器业务营收数额

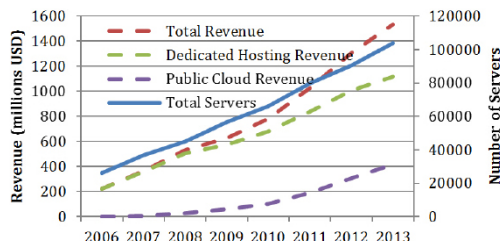


图6 Rackspace公司的服务器总数和营收数额

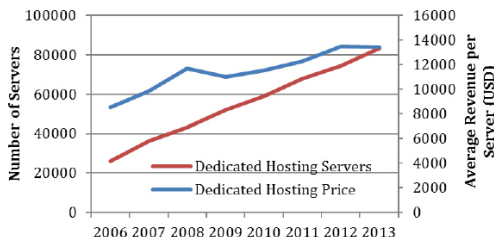


图7 Rackspace公司服务器租赁业务的服务器数量和平均价格

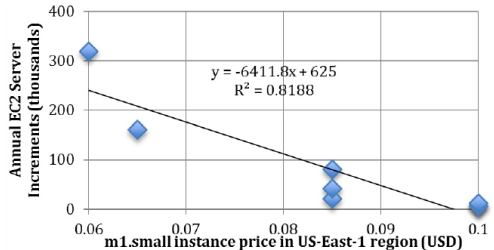


图8 Rackspace公司服务器租赁业务新增服务器数量和平均每台服务器营收

服务器租赁业务

图6为Rackspace公司2006到2013年间服务器数量和营收数额，包括服务器租赁业务营收和公有云服务营收。服务器总数以平均每年11000台的速度线性增长，而营收数额增长速度和服务器总数增长速度一致（两条直线的斜率基本一致）。自2009年起，服务器租赁业务营收增长速度明显放缓。2006到2009年间，公有云服务营收增长很缓慢，不到总营收的10%。2010年起，公有云服务营收增长较快。2013年公有云服务营收占全年总营收的27%。

为了研究服务器租赁业务的需求价格弹性，要估算Rackspace用于服务器租赁业务的服务器数量。为此，我们做了两个假设。其一，某项业务的营收与支持该项业务的服务器数量成正比，但在不同时期对不同业务要采用不同的参数；其二，随着公有云服务日渐成熟，平均每台服务器所贡献的营收要大于服务器租赁业务平均每台服务器所贡献的营收（不然，托管服务提供商完全没有必要投入资源研发公有云服务）。增加的营收主要来源为虚拟化和多租户所带来的效率提高。假设2008年以前公有云服务和服务器租赁业务的平均每台服务器营收是一致的，但2009到2013年间公有云服务的平均每台服务器营收从100%线性增长到150%（以当年服务器租赁业务的平均每台服务器营收为基准）。

图7是估算的Rackspace服务器租赁业务的服务器数量和平均每台服务器营收。在这段时间里服务器数量以每年7000台的速度线性增长，平均每台

服务器营收也呈线性增长的趋势（唯一的例外是2009年受全球经济危机影响，平均每台服务器营收略有回落）。总的来说，需求和价格同时增长，表明需求的刚性很强。对服务器销售业务来说，服务器出货数量代表的是当年新增加的计算机资源，且这些新增计算机资源的生命周期是三到五年。对服务器租赁业务来说，用于该业务的服务器数量是过去几年累积的。为了公平起见，我们使用每年新增的服务器数量来计算需求价格弹性，并将其与服务器销售业务进行比较。图8为服务器租赁业务的供需数据，横轴代表平均每台服务器营收，纵轴代表新增服务器数量。根据线性拟合结果，能算出2013年Rackspace公司服务器租赁业务的需

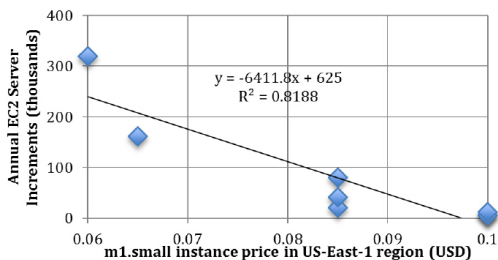


图9 EC2服务新增服务器数量和US-East-1区域m1.small按需实例价格求价格弹性为0.70，属于较强的刚性需求。

企业用户出于多种原因选择租赁服务器。对中小型企业来说，主要因为自建小型数据中心并维护所需要的网络链接并不合算，而租赁服务器，有多个服务提供商可供选择。企业通过筛选流程选择一家服务提供商，并与选定的服务提供商签订长期服务合同。合同跨度取决于企业对计算资源的需求，从几个月到几年不等。租赁服务器的企业对价格相对敏感，但长期合同的存在有效阻止了用户在市场价格发生波动时流向其他的服务提供商。

公有云业务

由于缺乏来源可靠的数据，分析AWS EC2的需求价格弹性相对困难。我们注意到区域US-East-1是最大的EC2区域，支撑了整个EC2大约50%的负载。此外，m1实例家族（上一代通用实例）是EC2上使用最多的实例类型，而m1实例的定价与其配置成正比。因此，可以基于以下几个假设对EC2的需求价格弹性进行估算。其一，m1.small按需实例的价格代表EC2服务正交分解后的价格；其二，EC2数据中心的利用率很高，因此EC2的服务器数量直接反映了对EC2服务的需求。与服务器租赁市场的分析相类似，我们感兴趣的是市场对新增计算资源的需求（用EC2的新增服务器来代表），而不是累积需求的总量（用EC2的服务器总数来代表）。图9为2008到2013年间EC2新增服务器估算值和m1.small按需实例的价格。根据线性拟合结果，能算出2013年AWS EC2业务的需求价格弹性为1.20，属于较强的柔性需求。

很多用户将EC2视为服务器租赁服务和虚拟私有服务器（VPS）的替代品，而忽略了云监控（CloudWatch）、自动伸缩（AutoScaling）、弹性负载均衡（Elastic Load Balancing）等高级特性。此外，花费在EC2上的开支（作为运营开支的一部

分）通常是在项目层面决定的，并且通常在项目总预算中占较大比例。可替代品的存在和预算比例很大共同决定了需求的柔性。也就是说，企业用户基于实际的业务需求决定其EC2开销，并且由于预算限制而对价格变化非常敏感。

通常来说，刚性需求与计划采购密切相关，而柔性需求则呈现冲动购物的特征。在固定资产模式下，企业在下单购买计算资源前首先需要申请必要的预算，而计算资源的交付和部署时间则长达数周甚至是数月。在云计算模式下，用户在一个可承受的计费框架下按需获取计算资源，避免了固定资产投资和计算资源交付周期。也就是说，云计算使得用户能够利用较小的预算完成较大的运算。因此，云计算能够催化出更多创新，并进一步创造出对计算资源的需求。

结论

基于2006到2013年的历史数据，本文研究了企业计算资源市场的需求价格弹性，主要结论如下。

- 服务器销售业务的需求刚性很强。企业根据其业务计划购买服务器，对价格变化并不敏感。IBM采取了正确的涨价策略，但市场的偏好正在从高端服务器往低端服务器迁移，导致了出货数量和营收数额的降低。HP采取了错误的降价策略，虽然出货数量有所增加，但营收数额反倒降低了。DELL采取了正确的涨价策略，又受益于市场偏好的迁移，同时实现了出货数量和营收数额增长。
- 服务器租赁业务的需求刚性较强。租赁服务器的企业对价格相对敏感，但长期合同的存在有效阻止了用户在市场价格发生波动时流向其他服务提供商。换句话说，企业根据业务规划签订服务器租赁合同，受市场上价格波动影响较小。
- 公有云服务的需求柔性较强。公有云用户对价格非常敏感，当价格下降时趋向于增加消费。P

本文是由悉尼大学信息技术学院蒋清野、Young Choon Lee和Albert Y. Zomaya共同撰写的，登载在本刊略有删减。
责任编辑：杨爽（yangshuang@csdn.net）