

浅谈“中国”语境下的公有云发展

文 / 蒋清野

公有云的规模

所谓公有云，简单地讲就是以服务的方式向公众提供计算资源。在这篇文章的范畴之内，计算资源主要指计算资源（虚拟机），但是在必要的时候会扩展到存储资源和网络资源。用各位从业人员背得滚瓜烂熟的术语来说，就是用户象用水用电一样使用计算资源，按需获取，按量计费。基于这样一个定义，一个真正意义上的公有云需要具备一定的规模才能够达到向“公众”提供服务的基本要求。“在这篇文章的范畴之内，托管云（Managed Cloud）被认为是公有云（Public Cloud）的一种特例。”

按照Gartner的统计数据，在2006到2014年间，全球服务器硬件市场每年的出货量稳定地在10,000,000台上下波动。其中，亚太地区占比在1/4左右，也就是2,500,000台。中国境内服务器出货数量在亚太地区的占比不详，保守地按1/5计算也有500,000台。按照3年折旧周期进行估算，全国范围内现役的计算资源至少有1,500,000台物理服务器。作为一家服务于“中国”的产业级别的公有云服务提供商，假设其业务成熟之后拥有全国计算资源的2%，就是30,000台物理服务器。再按1:3到1:4的虚拟化比例进行估算，则虚拟机的数量为100,000台左右。公有云作为一种新型服

务，其市场规模尚有相当程度的自然增长空间，因此五年之后的公有云可能达到的规模只会比这个数字更大。

根据AWS最近发布的财务数据，2015年第一季度的销售收入达到15.6亿美元。假设来自EC2以及基于EC2的其他服务对收入的贡献占50%，按照中等配置的m3.large实例（2个vCPU核心，7.5 GB内存，每小时0.14美元）来估算，相当于2,500,000个EC2实例。根据Rackspace历年的财报进行估算，2014年Rackspace用于公有云服务的物理服务器数量大概在20,000台到30,000台之间，换算成虚拟机也达到了100,000台。因此，将100,000台虚拟机作为一个基础目标，并非好高骛远。

基于这些估算，我们可以根据其规模判断一家公有云创业企业所处的成长阶段：

- 概念阶段，小于5,000台虚拟机。公司的终极目标相对模糊，在私有云解决方案提供商和公有云服务提供商之间摇摆不定。在战术层面，缺乏明确的技术路线图，产品形态相对原始并且没有明确的技术指标。

- 原型阶段，小于10,000台虚拟机。公司基本上将其终极目标定位为公有云服务提供商。由于公有云和私有云之间的巨大差异，必然要放弃私有云解决方案服

务提供商的身份。在战术层面，基本形成相对清晰的技术路线图，基础产品（云主机）基本定型，在宕机时间和产品性能方面均有明确的技术指标。在云主机的基础上，提供能够承担中低负载的负载均衡、数据库、缓存等等周边产品。

■ 成长阶段，小于50,000台虚拟机。基础产品（云主机）能够满足高性能计算的要求，同时发展出一系列模块化的周边产品。普通用户完全依靠云服务提供商所提供的不同模块即可自主创建大规模可伸缩型应用（无需云服务提供商进行干预）。

■ 成熟阶段，小于100,000台虚拟机。在技术方面，资源利用率开始提高，规模效应开始出现。在市场方面，客户忠诚度开始提高，马太效应开始出现。这标志着公司在公有云领域已经获得了较有份量的市场份额，其产品和技术获得了一个或者多个细分市场的广泛认可。

■ 产业阶段，大于100,000台虚拟机。只有进入这一阶段，才能够认为一个服务提供商已经站稳了脚跟，可以把公有云当作一个产业来做了。至于最后能够做多，一个好看国内的大环境，另外一个还得看公司自身的发展策略。

按照这样一个阶段划分，国内大部分公有云创业公司都还处于概念阶段，最多有一家创业公司已经进入原型阶段。阿里云不能够按照创业公司来看待，但是如果只统计其ECS部分的话，可能处于成长阶段的早期。我个人的估计，五年后公有云拥有的计算资源可能占全国计算资源的3%到5%。这意味着市场可以容纳一大一小两家进入产业阶段的公有云服务提供商，外加两到三家进入成长阶段或者成熟阶段的公有云服务提供商在一些细分市场里面深耕细作。

这也就是为什么我一直强调云计算是一片刚刚显现的蓝海。现在国内各家做公有云的公司杀得你死我活，

看起来似乎已经是一片血海。在我看来，这些不过都是假象。如果一家公有云创业企业没有这样的大局观，那么我只有一个建议：“认怂服输，割肉止损，是为美德。”

公有云的产品

作为一个公有云服务提供商，其产品形态必然是多种多样的。但是公有云要取得成功，就不能是私有数据中心可有可无的补充，而必须具备完全取代私有数据中心的能力。这意味这公有云要能够满足高性能计算的要求，普通用户完全依靠云服务提供商所提供的各种模块即可自主创建大规模可伸缩型应用（无需云服务提供商进行干预）。12306的查询部分迁移到阿里云勉强可以算是一个案例，问题在于这个迁移需要阿里云内部工程师的深度参与，因此不能算是一个好的案例。

鉴于产品的多样性，这里我们仅以块存储、负载均衡、自动伸缩为切入点谈谈公有云产品的特性。

块存储的磁盘IO指标，在从业人士当中是一个热门话题。相关讨论大都集中在云主机磁盘应该达到什么级别的IOPS或者是吞吐量，其实这些讨论所关注的点是完全错误的。对于公有云服务提供商来说，重要的不是云主机平均可以达到什么样的IO指标，而是如何根据客户的需求对整体IO能力进行分配。对于需要10个IOPS的低流量企业主页，为其提供100个IOPS是没有必要的。对于需要1000个IOPS的企业级应用，为其提供100个IOPS是远远不够的。套用云服务“按需获取，按量计费”的思路，IO能力需要成为可以“按需获取，按量计费”的商品。对于需要大容量低性能的用户，可以卖存储空间；对于需要低容量高性能的用户，可以卖IOPS。譬如说AWS提供三种不同规格的EBS卷：传统机械硬盘EBS卷（magnetic）不论磁盘大小平均

提供100个IOPS的IO能力，GP2型SSD EBS卷每一GB保证提供3个IOPS同时又可以允许高达3000个IOPS的爆发性IO，Provisioned IOPS型SSD EBS卷保证可以达到用户创建该EBS卷时所指定的IOPS指标。有了这样的设计，用户可以根据其实际需求购买所需要的磁盘空间或者是IOPS。尽管这样的购买依然受到服务提供商整体IO能力的限制，但是至少比所有的云主机都具备类似的“平均性能指标”要好得多。显而易见，设计这样的产品，要求云服务提供商对计算资源具有极细颗粒度的调控能力。

负载均衡也与此类似。对于一个正常的Web应用，其负载通常可以划分成三个档次：长期平均负载，长期高峰负载，短期爆发负载。在每秒只有数百个请求的情况下，负载均衡具备每秒处理一万个请求的能力是没有必要的。在每秒达到数万个请求的情况下，负载均衡只有每秒处理一万个请求的能力是远远不够的。如果用户按负载峰值购买负载均衡，结果是资源利用率偏低；如果用户按负载平均值购买负载均衡，结果是高峰期访问质量降低；如果用户按照实际负载切换负载均衡，结果是他再也不敢用公有云了。因此，负载均衡也要根据“按需获取，按量计费”的思路来设计，在负载降低的时候自动降级，在负载升高时自动升级。这样一种特性，就是自动伸缩。

将自动伸缩这个概念应用到云主机集群上，就是AWS的AutoScaling Group (ASG)。一个ASG包含一组具备相同功能的云主机，应用负载降低的时候，ASG自动杀掉多余的云主机以节省成本；应用负载升高的时候，ASG自动启动更多的云主机以应对压力。用户按照系统的实际负载购买计算资源，既不存在处理能力不足的问题，也不存在浪费计算资源的问题。

如上几个例子，都是AWS在其发展早期就已经实现的技术，其核心思想都是“按需获取，按量计费”。更重要的是，通过自动伸缩这样的概念，在满足客户负载

需求的前提下没有让客户花冤枉钱。我在前段时间写了一个题为《Building a scalable web application from ground zero》^[1] 的入门小教程，基本上能够反映一个中型Web应用对计算资源的需求特征。各位做公有云的不妨对照这个教程看看类似的需求如何在自己的平台上实现。AWS可能不是公有云的终极模式，但它至少是一种相对先进的模式，其产品对同行来说是极具启发意义的。一家公有云领域的创业公司，如果了解不熟悉AWS的产品，未免有闭门造车之嫌了。

有些人可能会说，AWS的产品好是好，但是国内用户并不接受。这就涉及到创业公司到底是想做现在的市场还是想做未来的市场的问题。如果做现在的市场，就必须迎合市场的需求，按照客户的要求去设计你的产品。如果做未来的市场，就必须从技术上进行创新，指导客户按照你的思路去设计他的应用。最近几年，国内市场（尤其是互联网公司）对AWS所倡导的理念的接受程度是在稳步提高的。对比国际上几家公有云服务提供商，目前的局势是AWS一家独大，剩下几家（包括Rackspace、Windows Azure、Google Compute Engine、HP Cloud）容量的总和与AWS存在接近一个数量级的差别。究其原因，在于其他几家出于种种原因没有接受AWS所倡导的“按需获取，按量计费”理念，只是按照传统数据中心的思路来做公有云而已。在这个大背景下，国内创业公司在熟悉AWS产品的基础上，模仿AWS的产品并争取有所创新，可能是创业早期（譬如说概念阶段）相对稳妥的发展道路。

公有云的成长

公有云的成长，涉及两个问题：一是用户增长，一是财务回报。

在用户增长方面，阿里云目前的方法有两个，一个是

将存量用户（万网的用户，天猫的商户）往云上迁移，另外一个发展政府客户。这两种客户，其特点都是对负载的要求不高（天猫整体的负载很高，但是大部分商家的独立负载并不高），对“按需获取，按量计费”的需求并不明显。换句话说，基本上是将公有云当作传统的服务器托管的替代品来用。以阿里云目前的状况来看，将这两部分用户做好只是时间问题。从规模上看，把这两部分用户做好了，阿里云应该可以从成熟阶段进入产业阶段。问题在于，做好这两部分用户只能让阿里云拥有公有云的皮毛，并不能让阿里云拥有公有云的本质。这种情况和Rackspace往公有云转型过程中所遇到的问题类似。Rackspace创立于1998年，以服务器租赁起家，平均每年新增服务器数量10,000台左右。受AWS的影响，Rackspace从2008年起开始做公有云，但是其思路一直是用虚拟机替代物理服务器，并没有从“按需获取，按量计费”这样的思路去设计其公有云产品。仔细研究Rackspace从2006年到2014年间的财报数据，可以看到其收入总额和服务器数量基本上呈线性增长的趋势。换句话说，Rackspace只是在做物理服务器的替代品，公有云部分并未对其业务产生重大影响。另外，一个值得探讨的问题是在“中国”这个语境下是否真的需要类似于AWS的“按需获取，按量计费”的公有云？或者说，“按需获取，按量计费”这样的需求，在所有需求中到底占多大份量。根据个人的观察，“按需获取，按量计费”这样的理念，即使是在国内互联网行业当中也还有待进一步推广，在其他行业中的接受程度显然要更低。受政策影响，未来三到五年政府在计算资源采购方面全面向公有云倾斜，而这部分用户关心的只是供应商的名字是否有“云”字，至于这个“云”字后面是啥完全不在考虑之列。我不止一次听在政府部门做IT的同行说领导要求项目一定要用上阿里云，至于用阿里云干啥完全没有要求。因此，每次有朋友问我

阿里云值不值得去的时候我都说阿里云的前景一片光明，如果能去的话当然要去。

按照王博士早些年的想法，阿里云还要为阿里巴巴集团提供服务。在王博士执掌阿里云的时期，阿里巴巴内部的人都觉得这是个笑话，不仅内心厌恶而且公开抵制。（关于王博士的故事，可以参考我两年前写的一篇文章《从王博士说起》^[2]。）现在章文嵩等人成为阿里云的主力，这个笑话便有了变成现实的可能性。至于这个可能性有多大，还得看阿里云后面两到三年的发展。一旦阿里云具备了为阿里巴巴集团提供服务的能力，为其他互联网企业提供服务更是不在话下。届时，阿里云可能会成为国内公有云领域毫无疑问的老大。2012年5月我在第四届中国云计算大会的一个演讲上说“阿里云的技术也很好，但是在云计算产品的设计方面，还是比较业余的”，当时在从业人员当中引起了很大争议。三年过去了，如果在同行内部做一个横向比较的话，阿里云的基础产品和某些创业公司的产品相比尚存在较大差距。这个差距并非来自技术差异而是来自认知差异——换句话说，不是因为阿里云的工程师们技术水平不行，而是因为阿里云还是没有从公有云的角度去设计产品。

与阿里云相比，创业公司基本上属于“三无”状态：没有存量用户、缺乏政府资源，尚未形成品牌。创业公司的用户增长过程，一期靠创始人的人品，二期靠技术推广，三期靠定向销售。所以创业公司的用户一般可以分成两类：某细分行业用户，其他创业公司。因此，创业公司更有可能根据自己的发展思路对其早期用户进行教育，指导早期用户按照自己的思路和产品路线设计应用。这些投入在公司发展早期看似无用，但当客户的业务逐步增长而公有云并不成为其负载或者性能瓶颈的时候，他们就会成为公有云的长期客户和成功案例。2009年Netflix全面转向AWS时业内几乎全是等着看笑话的，现在Netflix是运行在公有云

上的最大型应用，同时也是AWS最有说服力的技术传教士。公有云帮助客户应对负载波动问题，使得客户可以聚焦在其自身业务上。客户的成功自然而然地导致消费增加，而其示范效应还会带来更多的客户。这样日积月累，方能形成一个良性循环。从资源投放的角度来看，提供“按需获取，按量计费”的能力要求云服务提供商预留部分计算资源用来应对客户的爆发性需求。云服务提供商只有到了一定的规模，才能够准确地预测客户对计算资源的需求，从而将闲置的计算资源降低到财务可以接受的比例之下。换句话说，客户成功才有公有云的成功，规模壮大才有公有云的盈利。

前两天看到陈沙克近期的一篇文章《一个做了15年运维的老兵对公有云的深度剖析》^[3]，开篇就谈到2014年做公有云的几家创业公司是否盈利。问题在于公有云市场不是一个短期市场，而是一个未来十年尚有充分增长空间的市场。目前，中国的公有云市场尚属于发展早期，应该专注于产品研发和客户教育。一家公有云创业公司如果在概念阶段就实现了盈利，这种盈利很有可能是不可持续的。在这里我想澄清一个广为流传的故事，那就是“由于其电子商务业务存在大量闲置计算资源，亚马逊想到了通过零售的方式盘活这些闲置资源，并在其基础上研发了公有云服务”。这样的故事听起来虽然合理，却是完完全全的无中生有。之所以对此进行澄清，是想说明AWS在其发展的早期同样会遇到客户教育、市场培养、需求预测等问题。通过接近10年的努力，AWS基本上解决了这些问题，并在国际公有云市场上取得了一家独大的地位。由于缺乏历史数据，我们无从得知AWS是在第几年开始进入盈利状态的。但是从S3业务的指数增长曲线来看，AWS不大可能在第四年（2010年）末就实现盈利。

谈到财务回报，就不能不谈公有云的计费模式和定价

策略。在《从微观经济学看云计算发展》^[4]一文中，我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。这些分析表明，和传统的服务器销售和服务器租赁业务相比，公有云改变的不仅仅是计算资源的商业模式，它改变的是计算资源市场的供需关系。对于服务器销售和服务器租赁业务来说，客户的需求是刚性的。这意味着客户通常是根据其业务规划购买计算资源，对计算资源的价格波动并不敏感。对于公有云业务来说，客户的需求是柔性的。这意味着客户对计算资源的价格波动相对敏感，在价格下降时趋向于增加消费。对比AWS和Rackspace，可以发现只有AWS呈现这个特性，Rackspace的云计算业务并没有呈现这个特性。因此，我把客户的需求到底是刚性还是柔性作为区分虚拟机租赁和“按需获取，按量计费”的公有云的标准。如果你的客户的需求是刚性的，那么你只不过是在用传统数据中心的思路在做虚拟机租赁业务；如果你的客户的需求是柔性的，那么你就是在做“按需获取，按量计费”的公有云业务。从业务增长的角度来看，传统数据中心基本上是线性增长，而“按需获取，按量计费”的公有云业务是指数增长。

一种经济现象的出现，与其参与者的行为是密不可分的。换句话说，不能因为在AWS那里观察到了柔性需求，就断言在中国一定也会出现柔性需求。关于这一点，Rackspace和HP Cloud恐怕深有体会，因为到目前为止他们还没有观察到柔性需求。在中国，创业公司如果沿用传统数据中心的思想来做公有云，结果只能是产品同质化市场红海化。反之，如果围绕“按需获取，按量计费”这个理念去进行创新，开始的时候可能相对困难，但是只有坚持下去才有走进公有云这片蓝海的可能。

在外人看来，阿里云可以说是要钱有钱，要牛有牛，有战略有战术，是公众心目中的土豪型选手，唯一的缺憾在于五行缺（对云计算有深刻理解的）产品经

理。依靠阿里巴巴的品牌和万网的销售能力，目前阿里云在国内的规模最大。但是从互联网行业的角度来看，阿里云的用户体验较差。很多人可能会认为阿里巴巴的技术很好，用阿里云应该比较放心。问题在于阿里巴巴并不等同于阿里云，就如同Google并不等同于Google Compute Engine，微软也不等同于Windows Azure。在互联网行业中，技术人员对青云和UCloud的认可度更高。虽然两者都还处于概念阶段，但是从其产品和运营来看，比较符合我对公有云的理解。这两者当中，青云看来更为激进，大有后起居上的势头。UnitedStack由于全面拥抱OpenStack而广为人知，目前还在私有云解决方案提供商和公有云服务提供商这两个角色之间摇摆不定。私有云和公有云固然都很好，但是往深了做是截然不同的两个方向。创业公司需要聚焦，因此UnitedStack需要尽早在这两个角色之间做一个决断。如果决定往公有云服务提供商这个方向去做的话，建议抽空看看OpenStack外面的世界。

（插播一下广告，Rackspace和HP都在用OpenStack来做公有云，两者都处于比较尴尬的状态。国内用OpenStack来做公有云的创业公司不妨思考一下，用OpenStack做公有云到底还缺少什么。我个人的直觉

是用OpenStack做底子不是不行，但是光有这个底子肯定不行。）

[1] <http://www.qyjohn.net/?p=3720>

[2] <http://www.qyjohn.net/?p=3436>

[3] http://weibo.com/p/1001603836309159446956?suda_ref=www.qyjohn.net

[4] <http://www.qyjohn.net/?p=3615>

延伸阅读

Data Source on the Economics of Computing Resource Market (<http://www.qyjohn.net/?p=3507>)



蒋清野

IEEE高级会员，于1999年获得清华大学学士学位，2000年获得美国伊利诺大学香槟分校硕士学位，目前是悉尼大学信息技术学院的博士研究生。他的研究领域包括并行与高性能计算、云计算的市场与经济、开源社区。

责任编辑：陈秋歌 (chenqg@csdn.net)