

20221014-CVM库存情况和运营策略汇报-to-dowson

powellli, 2022-10-01

Page 0 引入

dowson和各位老板下午好，我是云产品一部计算产品中心的李力，今天给各位老板汇报一下CVM的库存情况以及因此带来的运营策略调整。

本次汇报主要内部是在数据上分析库存大量产生的原因和解决方案，这些解决方案逐渐沉淀为具体的精细化运营策略。

本次汇报的时间大约在20分钟到30分钟。

Page 1 端到端利用率的变化

按照去年麦肯锡评价利用率的指标：端到端利用率，我们绘制了一个从2016年开始的端到端利用率的走势图。

我们可以看到在2016到2019年的这三四年的时间，CVM的规模涨了13倍，即使是原来比较粗放式的运营策略，仍然能做到一个比较好的端到端利用率。

然后从2019年底开始，拼多多的大量退还，给我们带来了前所未有的挑战，它所造成的空置无法快速被其它用户的增量所填补，反而是在2020年底开始其它的大客户（快手、虎牙、多点、riot、作业帮等教育行业客户）也开始发生大量退还。

过去的高速增长掩盖了粗放式运营的问题，在这两年的利用率提升专项工作中，我们逐渐去精细化调整运营策略，大概可以分为这几类：计费策略、需求调度、资源采购、机型设计。

简单来说，在2020年以前，在市场竞争之下包年包月的计费策略被击穿，用户可以随买随退，将风险转嫁到云厂商；并且用户提的需求不管是否理智，我们都是以饱和式供应的方式提前准备，因为用户突增需求很多，资源大多数时候是不够用的；在机型设计上会倾向于尽可能提供灵活的多种选择，当然这块我们比友商还是克制得多。这些共同造成了资源运营效率低和退还风险大的问题。

2020年以后的策略在这几个方面逐渐调整成为了：我们在计费策略上增加并鼓励去使用能保障承诺消耗的模式，在资源供应上在运管和供应链的统筹之下，不仅打通了公有云和自研上云的资源池子，也在建设端到端的需求到资源的管理体系，在机型上也通过硬件和虚拟化技术的提升，进一步缩减机型数量。

Page 2 国内和海外的库存情况

这一页详细拆解了一下库存产生的原因和现在优化的情况。

首先，我们看国内，国内的库存绝大部分是由拼多多等大客户退还和跳票导致的，在运管的帮助之下，早期公司帮助消化了大部分拼多多的空置设备，后来逐渐也通过EKS和Lighthouse分别在公司内外消耗剩余的旧代次碎片。可以说国内目前的库存情况已经有了非常好的改善。

国外的情况就要严峻得多，国外的库存大部分是由于快手和海外游戏的客户退还和跳票导致的。与国内不同的地方在于，海外很难找到类似于需求方，这也导致了海外库存消耗较慢。未来我们会通过海外搬迁来减缓这里的问题。

海外的库存问题较难解决还有一个原因是，事实上我们在海外的运营向来是比较谨慎的，海外主要是售卖标准型的。但是这些客户需求大部分反而是GPU、计算型等非标准机型，这也导致了这些机型闲置下来很难找到需求方。

我们最近跟TCI的同事也交流过一点，从数据上看，海外有很多非标准型的设备的售卖率，看起来是非标准型不好卖，而这其实只是一个数据结果，真实原因是客户要求采购然后退还的这些非标准型没有复用路径。

Page 3 库存占成本的比例变化

前面讲到了通过精细化的运营策略，库存从峰值降低了不少。这里有另外一个数据，从成本占比的角度上看，我们也可以看到，到今天，库存占成本的比例从2021年的11.3%下降到了5.5%。还是有一个比较好的数据表现。2021年35元，2022年8月30元左右。

除了库存以外，功耗优化等能力也进一步降低了成本。

单核成本差异集中在服务器折旧、资金成本及经营成本

CVM单核成本	阿里21年均	腾讯21年均	腾讯22年8月	21年均T vs A	22年8月T vs A
基础设施	16.58	22.63	19.54	6.05	2.96
服务器	8.26	13.07	10.38	4.81	2.12 ①
机架	7.09	7.13	7.13	0.04	0.04
网络	1.05	1.16	0.96	0.11	-0.09
外网IP	0.18	0.21	0.19	0.03	0.01
资金成本	0	1.06	0.88	1.06	0.88 ②
运营支撑	1.67	2.06	1.68	0.39	0.01
产品库存	2.85	5.22	2.64	2.37	-0.21
设备库存	1.54	3.88	1.49	2.34	-0.05
机位库存	1.31	1.34	1.15	0.03	-0.16 ③
公共分摊	0.24	0.62	0.37	0.38	0.13
经营成本	1.72	3.74	2.99	2.02	1.27 ④
小计	23.06	34.27	27.22	11.21	4.16
IEG结算及内领优惠	/	0.66	2.16		
合计		34.93	29.38		

① 服务器：过折旧期设备权重及Rebate差异影响

a、服务器采购：AMD占比从22%提升至29%
b、过折旧期设备占比：22年8月T-16% vs A-30%
c、Rebate: General Rebate按机型退还产品 vs 所有Rebate (含阶梯及其他特殊) 及BS价差穿透回机型

差异	21年均T vs 21年均A	22年8月T vs 21年均A
服务器采购	0.45	-0.05
服务器折旧	3.84	2.01
服务器Rebate	0.51	0.17

② 资金成本：

阿里对业务单元未核算资金成本
腾讯按集团财务口径核算硬件年化7%资金成本

③ 库存分摊：

库存成本持续优化后，22年H2持平
分摊成本因阿里云规模是腾讯云的2倍多，目标22年底优化后持平

④ 经营成本：差异主要在TEG服务费

差异	21年均T vs 21年均A	22年8月T vs 21年均A
TEG人力服务费	1.45	1.14
渠道成本	0.55	0.09
短信成本	0.03	0.03

注：虚拟化软件消耗合并到基础设施口径，22年8月IEG结算0.62元/核、内领优惠1.54元/核

1

Page 4 计费模式-锁定长期用量

接下来我们分别讲一下几个策略调整的方式的预期结果。

第一个是计费模式的升级。最早包年包月本来是一种挺好并且很简单的承诺用量的计费模式，但是从2017年开始，阿里云率先做了无条件的随时退还，逼得国内所有云厂商都逐渐跟进，从而导致包年包月实际变成了按月付费并可随时退还的模式。

我们近两周也了解到一个情况，阿里云运营团队内部也在复盘这个决策，它带来的结果是互卷和损人不利己。但是现在又很难在包年包月的计费模式上把这个约束条件重新补回来。

因此，我们在包年包月的基础上，针对那些用量特别大、异构机型或者是折扣特别低的场景下，需要在合同上让用户承诺包销，我们在产品上也设置了禁止提前退还。另外，更新的计费模式比如预留实例和节省计划也在推进当中。

当然，这块的比例目前还是非常少的，相对乐观的判断是，随着友商的策略调整，我们在市场竞争中的压力会小很多。

Page 5 需求调度-对外

下一部分是需求调度。

首先是对外的需求调度。以前在运营上我们倾向于让用户自己选择机型，那么用户总是会选最好最新的设备，这导致我们的旧代次和碎片很难消耗出去。今年我们针对库存现状，做了一些运营策略，引导更多的用户需求到闲置的资源上来。

目前已经消耗了200万核。

这个机制还需要持续增强

Page 6 需求调度-对内

对内是通过EKS来解决。自研上云的新阶段。

真正发生云原生的力量。

Page 7 跨地跨调度

不仅有内外部的资源池打通，也有跨地域的调度能力建设。

2022年年初至今搬迁超2.2万次，盘活资源超240万核

未来甚至可以盘活全球资源。

Page 8 供应链能力提升

如ppt

Page 9 端到端需求管理

拉通从行业到供应链的整个链条，确保用户需求到资源的全面管理，从BSC、预算、库存系统等多个系统的数据互通。

以上的库存促销、跨地域调度等也是统一在端到端需求管理的框架里面的。

端到端的需求管理是一个非常庞大和丰富的体系。

Page 10 机型收敛

过去，我们在机型设计上就比友商保守，每一代次大约只是友商的一半的机型。比如阿里云的标准机型，1:2、1:4、1:8分别是由三个母机型来承载，我们只会分为两个。

而在机型散乱的情况下，需求的管理会变得更复杂。也是得益于技术的提升，我们有机会进一步缩减母机型。

Ending

进一步精细化运营，进一步端到端管理，确保提升资源利用效率，降本增效。