

[亚马逊AWS官方博客](#)

你需要知道的关于高IO EC2的事儿

by [AWS Team](#) | on 07 SEP 2017 | [Permalink](#) | [Share](#)

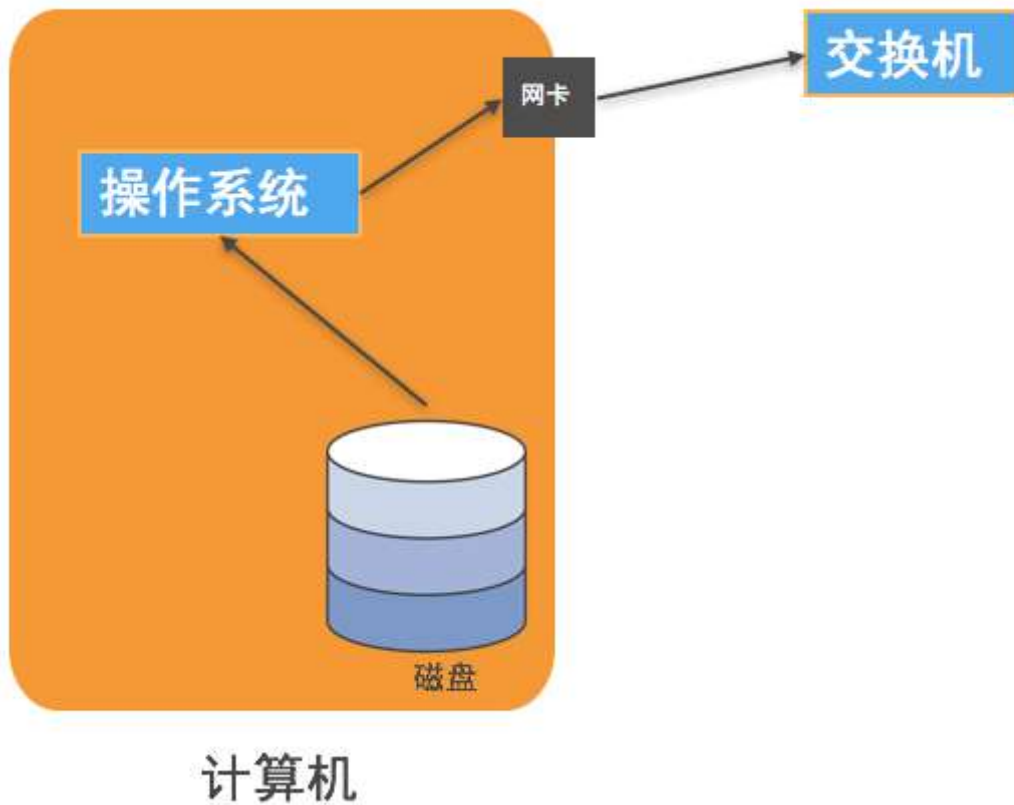
作者: 焦杨

故事背景

笔者长期在AWS从事架构师工作，人送焦导的外号。经常遇到客户抱怨：“我选了比原来大的机器，比原来快的硬盘，可EC2的IO怎么就是上不去，到底卡在哪里了啊？你们架构师到底怎么干活的啊？”，好了，今天我也不想再背这个锅了，我们一起把这个事儿好好说道说道。

基本原理

一开始，我们先来看看磁盘上的数据是怎么到达外网的。

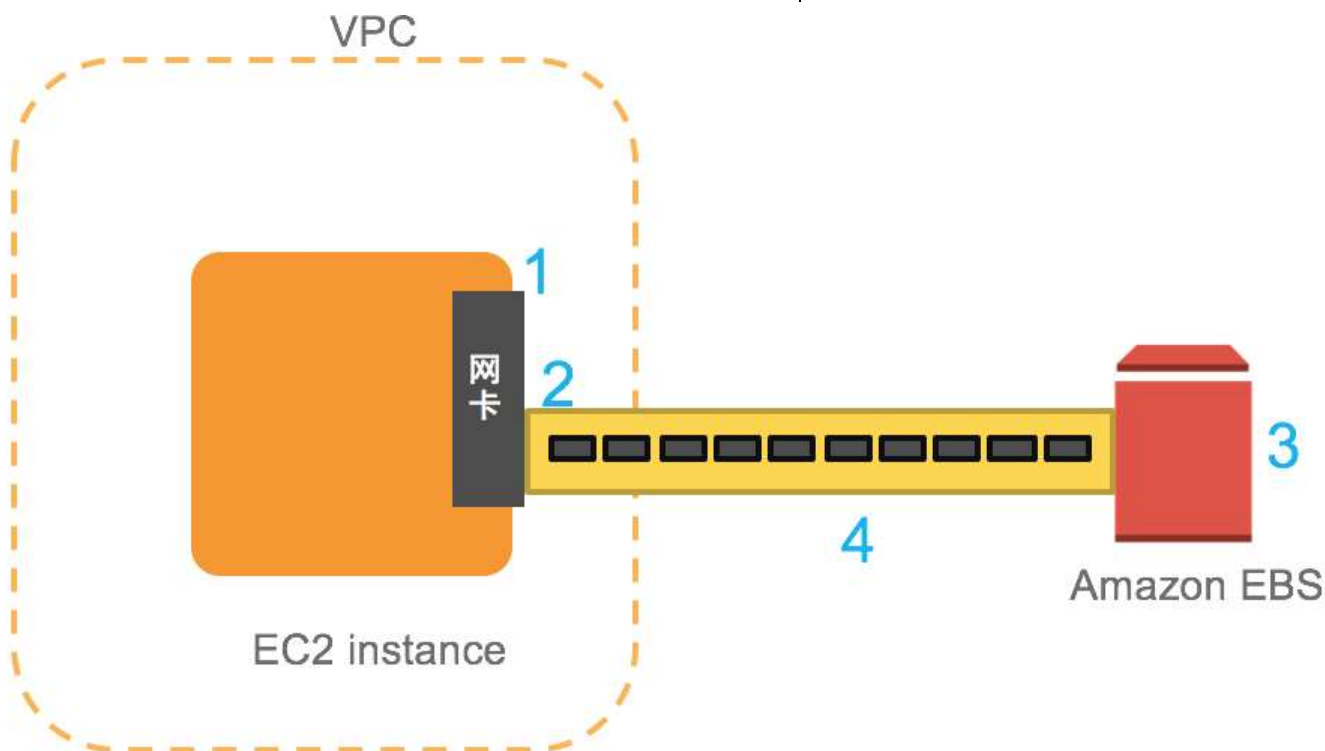


第一步操作系统接到IO命令，然后驱动磁盘，从磁盘上读入数据到内存处理，最后从网卡送出，通过交换机路由器送出到外部网络。要更好的IO性能，硬件层面上看无非三板斧

- 换个更块的硬盘
- 换个更快的网卡和交换机
- 更快的CPU、足够的内存、足够优化的

那到了AWS上，又变成什么样了呢？

- 云上的图



你可能看出来了：

- 物理机变成了EC2 instance。
- 磁盘为了更灵活和更可靠，把它从机器里拿出来变成了EBS。
- 外部的交换机/路由器由于VPC的存在，对用户变成透明的了。

稍微想想，发现实质完全没有变化。那是不是我们就可以用原有的三板斧解决性能问题了呢？

焦导的回答是：“完全没错，但是不够”。

大家知道，公有云服务商为了保证服务质量，避免所谓的“吵闹的邻居”问题，也为了避免意外操作导致的超额费用，引入了大量的QoS特性。这里边既有我们熟悉的各个service的hard和soft的limit，也有对外暴露的各种配置选项。为了要实现高IO的目的，我们有必要对这些QoS特性有清晰的了解。

还是上边的那个图，把其中的关键部分标上号，下边一一说明

列举分析

1. EC2网络

AWS向客户提供SDN的VPC网络，屏蔽了交换机、路由器等网络基础设施。但是对单个EC2实例的网络出口带宽是有明确的限制的。

实例类型	默认情况下为 EBS 优化	最大带宽 (Mbps)*	预期吞吐量 (MB/s)**	最大 IOPS (I/O 大小为 16KB)**
c1.xlarge		1000	125	8000
c3.xlarge		500	62.5	4000
c3.2xlarge		1000	125	8000
c3.4xlarge		2000	250	16000
c4.large	是	500	62.5	4000
c4.xlarge	是	750	93.75	6000
c4.2xlarge	是	1000	125	8000
c4.4xlarge	是	2000	250	16000

你可能会说这不就是网卡限制吗？多加几个网卡不就行了。

请注意，增加多个网卡并不能增加单个EC2的总出口带宽。

更加详细的数据请查看这里http://docs.aws.amazon.com/zh_cn/AWSEC2/latest/UserGuide/ebs-ec2-config.html

2 存储网络

默认情况下，刚才提到的EC2总带宽，包含了进出EC2实例的所有流量。熟悉网络的同学应该知道，这里边包含了EC2间的计算流量、访问存储设备的存储流量等的各种流量。

要优化性能，拿出单独的网络来走存储流量就可以了。这也就是我们通常说的计算、存储流量分离技术。在AWS里，我们把这个叫做“EBS优化”。

实例类型	默认情况下为 EBS 优化	最大带宽 (Mbps)*	预期吞吐量 (MB/s)**	最大 IOPS (I/O 大小为 16KB)**
c1.xlarge		1000	125	8000
c3.xlarge		500	62.5	4000
c3.2xlarge		1000	125	8000
c3.4xlarge		2000	250	16000
c4.large	是	500	62.5	4000
c4.xlarge	是	750	93.75	6000
c4.2xlarge	是	1000	125	8000
c4.4xlarge	是	2000	250	16000
c4.8xlarge	是	4000	500	32000
d2.xlarge	是	750	93.75	6000
d2.2xlarge	是	1000	125	8000

当然了，对于一些早期机型（i2,c3等），在使用了万兆网络的情况下。因为这根管子已经足够粗，流量混在一起问题也不是很大。

EC2机型EBS优化的更多信息，可以在这里找到

<https://aws.amazon.com/cn/blogs/china/what-you-need-to-know-about-high-io-ec2/>

http://docs.aws.amazon.com/zh_cn/AWSEC2/latest/UserGuide/ebs-ec2-config.html

3 磁盘性能

要想获得高IO，选择合适的磁盘应该是最基本的了吧。是看重IOPS而选择SSD型，还是看重连续读写的磁盘吞吐而选择HDD，这是你要做出的关于磁盘的第一个决定。

	固态硬盘 (SSD)		硬盘 (HDD)	
卷类型	通用型 SSD (gp2)*	预配置 IOPS SSD (io1)	吞吐优化 HDD (st1)	Cold HDD (sc1)
卷大小	1 GiB – 16 TiB	4 GiB – 16 TiB	500 GiB – 16 TiB	500 GiB – 16 TiB
最大 IOPS**/卷	10000	20000	500	250
最大吞吐量/卷†	160 MiB/s	320 MiB/s	500 MiB/s	250 MiB/s
最大 IOPS/实例	75000	75000	75000	75000
最大吞吐量/实例	1,750 MB/s	1,750 MB/s	1,750 MB/s	1,750 MB/s

这里要特别注意的几个点：

- 单卷IOPS的具体值和容量正相关，对于GP2而言，基准值为3IOPS/GB
- 对单卷有IOPS和吞吐量的限制，两个因素同时起作用
- 对连接了多个卷的实例，总体IOPS和吞吐量也分别有限制

另外需要注意的是，对于使用最多的GP2类型SSD，存在着读写突增特性（链接），无论多小的磁盘短时间内IOPS都可以到达3000，具体就不再展开说了。参考这里

http://docs.aws.amazon.com/zh_cn/AWSEC2/latest/UserGuide/EBSVolumeTypes.html

4 存储队列

卷存储队列是指等待设备处理的 I/O 请求的队列，它的长度表明了有多少I/O请求还没有得到执行。队列过长，读写延迟加大，队列过短，读写任务不饱满。要让EBS全速工作起来，需要有足够的任务，也就是说要让EBS始终处于忙碌的状态。

基于这个认识，有必要根据工作负载的具体状况，实时观察存储队列长度的情况。对工作负载进行适当调整，以便发挥底层设备的最大能力。

这个指标可以通过Cloudwatch的VolumeQueueLength 来观察。

http://docs.aws.amazon.com/zh_cn/AWSEC2/latest/UserGuide/ebs-io-characteristics.html

综合和特例

还有一些特殊的情况也值得拿出来说一下：如果我们特别在意存储网络的带宽的限制，完全可以使用带有instance-storage的实例类型，比如i系列和d系列。存储在本地，通过本地总线连接，就没有这个存储网络的限制啦。

某些情况下，追求极致的IOPS，也可以考虑使用条带化技术将多个卷合起来一起使用。

另外，如果我们的流量模型为从磁盘上一次读出，在计算节点上向多个client分发，简单的说就是具有fan-out效果的话，碰到瓶颈的地方大概率应该在EC2 instance的流出限制之上，而不是存储网络限制。

总结

总结一下，要获得期望的高IO。你需要准备

- a. 合适的磁盘类型，磁盘个数，保证你需要达到IO在单卷和总卷限制范围以内。

- b. 足够快的从磁盘到EC2的高速网络
- c. 足够大的机型，以带来足够的inbound/outbound。
- d. 适当的存储队列，保证磁盘够忙

好了，需要知道的所有关于EC2的IO性能的秘籍都在这儿了，马上动手，去榨干你购买的EC2的所有潜能吧。

作者介绍：



焦杨，AWS解决方案架构师，IT行业老兵，在Moto，NEC，路透，汉柏，AWS等国内外企业从业10余年。2011年之前，从事过web container, EJB container 和大型网站的后台研发工作，醉心于OO和设计模式。2011开始着手基于oVirt和Openstack的私有云平台研发工作，并亲密接触SDN，关注点随之转移到基础架构、分布式和高可用，自此自诩全栈工程师。2015年加入AWS，担任解决方案架构师，负责基于AWS的云计算方案架构的咨询和设计工作。

TAGS: [大咖专栏](#), [计算](#)

Related Posts

[AWS 官方博客目录](#)

[使用 Amazon API Gateway 为 SAP 部署 API](#)

[SAP on AWS – 过去、现在和未来](#)

[Amazon Aurora 在西云数据运营的 AWS 中国（宁夏）区域落地](#)

[将 MySQL 数据库迁移到 Amazon Aurora 数据库](#)

[基于 Amazon ECS 的并行批处理任务解决方案](#)

[AWS 如何帮助客户及 APN 合作伙伴推进 GDPR 合规进程？](#)

[建立 VPC 并基于参数动态创建子网的 CloudFormation 模板](#)

