[http://cdn1.infoq.com/styles/cn/i/icon-right.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html) 余锋先生您好。欢迎您参加QCon0：04并给我们做精彩的演讲。也谢谢您接受我们的采访。您能向观众朋友大概介绍一下您自己，包括您现在负责的具体工作是什么吗？ [http://cdn2.infoq.com/styles/cn/i/icon-collapse.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html)

大家好，我隶属于淘宝网核心系统数据库组。我们的团队在淘宝主要负责MySQL分支的维护。大家知道淘宝去年开始在做去Oracle化的工作，由于涉及到历史数据保留，替代方案当时选择的是MySQL。由于淘宝很多的产品都用到了数据库，这也促使MySQL即将会慢慢被大规模使用。Oracle是一个商业的产品，所以它有很多的特性是非常适合大规模数据使用的。同时MySQL作为一个开源的产品，也具有很多商业数据库所没有的特性，于是我们就要去填补这个空白。我们的工作就是在开源的MySQL基础上做一些改进和优化，特别是性能方面的优化。MySQL最初的设计不是用来存储大规模数据的，但淘宝的数据量非常惊人，所以在I/O方面，尤其是CPU I/O层面会有很大瓶颈，因此我们的主要目标也是解决IO方面的瓶颈问题

[http://cdn3.infoq.com/styles/cn/i/icon-right.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html) 在谈到具体的MySQL优化策略之前，我想大家都比较好奇，介于淘宝现在的数据容量，每天的吞吐量是一个什么样的情况？ [http://cdn2.infoq.com/styles/cn/i/icon-collapse.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html)

淘宝的产品线很多，以我们目前接触到的淘宝的商品库为例，淘宝商品库是什么概念？就是卖家卖的所有的商品都保存在数据库上。我们都知道淘宝商品是非常多的，有可能是十亿级别的，而且淘宝每年的数据都会翻番，过去亦是如此，未来也会如此，动辄几十亿条的记录。在淘宝上，由于卖家是非常活跃的，所以每天的交易数也是非常多的，而且每天的访问量也是相当大的。更比如在一些高峰期，比如节假日，像去年的四个一光棍节促销，当时一天的访问量可就是平常的好几倍，几十倍的样子。具体的数字，在网站上应该会看到，这个数字也在不停的变化，大概会在10亿的这么一个数量级上。

[http://cdn3.infoq.com/styles/cn/i/icon-right.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html) 在Oracle 切换到MySQL后，可能会有一些商业上的特性在MySQL上无法得到体现。所以在做优化之前，是否设定了一些策略的标准，比如说一定要达到什么样的标准才算是成功的优化？ [http://cdn2.infoq.com/styles/cn/i/icon-collapse.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html)

我们的标准主要基于以下三个方面，首先是安全性。因为淘宝卖家的商品都存在上面，一大堆数据，商品没了，淘宝就会很难受；其次是性能，因为数据量大，而且未来增长的速度会非常快，以前我们的系统是跑在Oracle的小机上面，它的成本是非常高的。现在改用高性能的PC服务器，性价比较高，而且Oracle数据库的能力不足以维持淘宝的更新速度；最后是可运维性，Oracle是有一套完整的工具，利用它们我们可随时查看相关的状态和预测它的一些变化。MySQL作为开源产品，这部分的功能稍显欠缺，因此我们就需要根据运维团队和DBA的需求去开发一些工具。

[http://cdn2.infoq.com/styles/cn/i/icon-right.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html) 就是在日常优化工作的过程中帮助你去运维？ [http://cdn2.infoq.com/styles/cn/i/icon-collapse.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html)

对，这也算优化的一方面，算过程的一个方面。

[http://cdn1.infoq.com/styles/cn/i/icon-right.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html) 在产品的优化过程中，是否会发现用MySQL替代原来的Oracle有些问题还是无法避免的，您是如何处理的？ [http://cdn2.infoq.com/styles/cn/i/icon-collapse.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html)

是的，是有这样的问题，比如Oracle的双机备份，这块就做得非常好，整体的方案都不错。相比之下MySQL这点就比较弱，我们必须在过程中去深化解决这个问题。还有像Oracle使用的高端存储，它的高端存储的I/O能力是非常强的。现在我们改用PC服务器，如果用硬盘来做的话，I/O能力会很差，甚至说I/O的能力在一两千I/O PS就已经算很高的了，这样的话相比原来的高端存储会相差一个数量级。我们在过去几个月顺利地解决了这个问题，而且结果证明，性能要比高端存储还要好。

[http://cdn1.infoq.com/styles/cn/i/icon-right.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html) 那么这个优化的工作现在是一个什么样的状态，进行中还是已经上线？ [http://cdn2.infoq.com/styles/cn/i/icon-collapse.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html)

从去年8月开始到现在已经持续了好几个月，这个月就可以上线。其实主要的攻坚工作在两个月前就已经做好了，由于要采购设备，所以安排在这月上线，上线后即可替代原来的Oracle数据库。

[http://cdn3.infoq.com/styles/cn/i/icon-right.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html) 就目前的结果来看优化成果如何？ [http://cdn2.infoq.com/styles/cn/i/icon-collapse.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html)

从机器的运算能力上来看的话，比原来要高很多，从设备的总体的价格来看的话，相比原来小机的价格，现在的价格相当于原来的几分之一。说的通俗一点就是，花更少的钱做更多的事情，基本是这样的一个情况。

[http://cdn2.infoq.com/styles/cn/i/icon-right.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html) 优化MySQL的过程中有没有涉及到水平分片，或者垂直分片，大概的策略是什么？ [http://cdn2.infoq.com/styles/cn/i/icon-collapse.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html)

这块因为之前的MySQL从业务的角度上我们已经优化过了。涉及到业务上的复杂SQL查询，基本上都把它当成Key-Value来处理。所以说在水平切割上就变得非常简单，我们只需要根据用户的这些ID把它平均地分割到集群的各个机器上去即可，所以说这块会做得非常简单，也不会有那些非常复杂的SQL。

[http://cdn3.infoq.com/styles/cn/i/icon-right.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html) 刚才提到在去Oracle化过程中，最终选择了MySQL。是怎样的原因使你没有去选择NoSQL或是其它技术来作为提高性能优化方面的考虑呢？ [http://cdn2.infoq.com/styles/cn/i/icon-collapse.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html)

主要考虑到以下这几个方面，第一是因为MySQL已经是非常成熟的产品，而且也是经过大规模系统考验的，所以它的稳定性，开发的便捷性还是比较有优势的。从另外一个角度上来讲，也能最大限度的保障历史系统的迁移，这是第一个方面。此外，之所以没有采用NoSQL主要基于以下几个方面的考虑，第一由于市面上流行的NoSQL方案相对来说都比较年轻。然而对于淘宝商品库来讲，它是作为淘宝非常重要的库，是不允许我们有丝毫闪失的，不能出现任何问题。第二点，现在流行的NoSQL方案，从程序的架构上来讲，就是对传统数据库的过程进行了精简。但是它的存储、内存等问题其实还是存在的。只是解决问题的思路变得轻量。但是这些问题依旧在的。 现在我们在数据库层面，因为平常都是从引擎层面去开发，在原码级别去做优化，甚至包括操作系统级的优化。所以我们可以很清楚的知道问题在哪里，应该解决什么问题，最终能达到什么样的效果。那这样做的好处就是我们既能解决问题，又能充分利用传统数据库的方便性以及稳定性。

[http://cdn4.infoq.com/styles/cn/i/icon-right.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html) 请您分享一下，在数据库优化方面，尤其是MySQL优化这部分，在优化的过程种需要遵循那些原则？ [http://cdn2.infoq.com/styles/cn/i/icon-collapse.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html)

优化听起来好像没什么内容，但实际上我们做了很多的工作。首先要有明确的目标，我们的目标是什么，是为了提高性能还是提高安全性，还是提高其他的什么指标，这个目标要清晰和完整。第二点是测量，你不能说，我拍脑袋看到某某某，他怎么优化的，然后我也上去这样照着优化，这不是我们做事的风格。我们要测量现在的瓶颈在哪里？我们的系统遇到了什么问题，这点我们会借助一些工具来实现。比如说测量CPU的使用，我们会非常精确的去测量I/O的使用，比如说在设备层面，I/O是怎么使用的，在软件系统层面I/O是怎么使用的，在数据库引擎层面I/O是怎么使用的。通过非常精细的测量，我们就能找出，对I/O使用不恰当的地方，然后就是要去解决这种问题，为什么会比设想的要多呢，或者为什么要比设想的要慢呢，我们会针对此类问题做理论上的分析，甚至会去翻翻源码它到底是怎么实现的，这样我们就可以比较有针对性将问题解决。需要特别提到的是，我们这次有一个创新，用高速的SSD盘去做2级Cache，因为传统的MySQL数据库都是引擎里面发挥pool的功能，然后直接在软件系统存储。就是引擎里边吐出来东西到那边去，命不中东西从磁盘读进去，由于是传统的磁盘，所以它的寻道时间是非常长的。那它的I/O其实是不高的，但是吞吐量大。为了解决问题，我们采用一种叫PCI-E的Flash卡。这个卡的特点是，它是电子盘，所以它没有寻道时间，随机特性非常好，同时它的读写吞吐量非常高，读能大概到1G，写能达到几百兆。采用其作为二级cache，数据从InnoDB引擎这边出来，不到磁盘去，而是先写到高速卡中。由于高速卡很快，它完成的I/O PS是微秒级的，几十微秒就写完了，然后传统磁盘却是十个毫秒级的。所以从数据库引擎的角度来看，只需写入几十微妙就解决了。写入数据开始是随机的，随机累计起来以后它就不是随机的，我们通过把它排序，就可以使其变为顺序的。然后我们再把这顺序的数据利用磁盘的高吞吐量特点，由于数据库半夜时候它可能会比较轻松。那这时候，把它导到磁盘中去，这样就大大的提高整个系统的IO能力。这是比较大的一个创新点。

[http://cdn1.infoq.com/styles/cn/i/icon-right.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html) 那我们对您的采访就到这里结束了，谢谢您。 [http://cdn2.infoq.com/styles/cn/i/icon-collapse.gif](http://www.cnblogs.com/end/archive/2011/05/20/2052462.html)

谢谢。