

知 某局点进行ONESTor性能测试，测试小IO顺序写，在测试一段时间后性能下降问题分析

yucan 2020-01-17 发表

组网及说明

某局点ONESTor 3.0独立部署，flashcache回写模式缓存加速。划分存储卷给VMware客户端上的redhat虚拟机使用，测试小IO顺序写，当存储卷容量写到70%时，测试性能会下降一半左右。

问题描述

从虚拟机内部测试小IO顺序写，当存储卷容量写到70%时，测试性能会下降一半左右。

过程分析

ONESTor使用SSD flashcache作为缓存加速，由于SSD对于小IO的处理比HDD好得多，对于4K随机写的测试，测试IO会先落到flashcache SSD中，然后就返回落盘，以实现写加速的作用。

落到flashcache盘里的数据，不会立即下刷到HDD，而是留在flashcache中，这样下次如果要读这部分数据，就可以直接从缓存中读取，以实现读加速功能。

当flashcache盘快要被写满的时候，此时会通过一种淘汰算法，主要是依据数据访问的热度，来决定哪部分数据是热数据，哪部分数据是冷数据，热数据留在flashcache中，冷数据下刷到HDD，以腾出空间给新的热点数据。

进行性能测试时，如果采取小IO顺序写，相当于所有的数据都只访问一次，数据没有冷热的区别。这样就会使得flashcache中的数据不会被淘汰，也就不会下刷。当存储卷写到70%左右的时候，缓存就几乎写满，此后下发的IO会直接写HDD，相当于没有缓存，因此性能会下降。

在真实的业务环境下，不会一直都是小IO新数据写入，不管是读取还是修改，总会去访问到以前的旧数据。如果是修改热点数据，就会直接去写SSD里存有该热点数据的区域，会比写HDD快得多。因此真实情况下，反而不会出现磁盘将近写满时的性能大幅下降的现象。

解决方法

在进行性能测试时，如果要测磁盘在小IO下的IOPS，业界的通用做法是测试小IO下的随机读写。如果要测试带宽，则应该测试大IO下的顺序读写。