

# linux cache介绍及控制



lihanglucien [关注](#) IP属地: 湖南

0.096 2019.03.27 13:57:59 字数 1,119 阅读 3,088

## 参考文档

[https://lonesysadmin.net/2013/12/22/better-linux-disk-caching-performance-vm-dirty\\_ratio/](https://lonesysadmin.net/2013/12/22/better-linux-disk-caching-performance-vm-dirty_ratio/)

## 有关Cache

文件缓存是提升性能的重要手段。毋庸置疑，读缓存（Read caching）在绝大多数情况下是有益无害的（程序可以直接从RAM中读取数据），而写缓存（Write caching）则相对复杂。Linux内核将写磁盘的操作分解成了，先写缓存，每隔一段时间再异步地将缓存写入磁盘。这提升了IO读写的速度，但存在一定风险。数据没有及时写入磁盘，所以存在数据丢失的风险。

同样，也存在cache被写爆的情况。还可能出现一次性往磁盘写入过多数据，以致使系统卡顿。之所以卡顿，是因为系统认为，缓存太大用异步的方式来不及把它们都写进磁盘，于是切换到同步的方式写入。（异步，即写入的同时进程能正常运行；同步，即写完之前其他进程不能工作）。

好消息是，你可以根据实际情况，对写缓存进行配置。

可以看一下这些参数：

```
[root@host ~]# sysctl -a | grep dirty
vm.dirty_background_ratio = 10
vm.dirty_background_bytes = 0
vm.dirty_ratio = 20
vm.dirty_bytes = 0
vm.dirty_writeback_centisecs = 500
vm.dirty_expire_centisecs = 3000
```

**vm.dirty\_background\_ratio** 是内存可以填充“脏数据”的百分比。这些“脏数据”在稍后是会写入磁盘的，`pdflush/flush/kdmflush`这些后台进程会稍后清理脏数据。举一个例子，我有32G内存，那么有3.2G的内存可以待着内存里，超过3.2G的话就会有后来进程来清理它。

**vm.dirty\_ratio** 是绝对的脏数据限制，内存里的脏数据百分比不能超过这个值。如果脏数据超过这个数量，新的IO请求将会被阻挡，直到脏数据被写进磁盘。这是造成IO卡顿的重要原因，但这也是保

证内存中不会存在过量脏数据的保护机制。

**vm.dirty\_expire\_centisecs** 指定脏数据能存活的时间。在这里它的值是30秒。当 `pdflush/flush/kdmflush` 进行起来时，它会检查是否有数据超过这个时限，如果有则会把它异步地写到磁盘中。毕竟数据在内存里待太久也会有丢失风险。

**vm.dirty\_writeback\_centisecs** 指定多长时间 `pdflush/flush/kdmflush` 这些进程会起来一次。

可以通过下面方式看内存中有多少脏数据：

```
[root@host ~]# cat /proc/vmstat | egrep "dirty|writeback"
nr_dirty 69
nr_writeback 0
nr_writeback_temp 0
```

这说明了，我有69页的脏数据要写到磁盘里。

## 情景1：减少Cache

你可以针对要做的事情，来制定一个合适的值。

在一些情况下，我们有快速的磁盘子系统，它们有自带的带备用电池的NVRAM caches，这时候把数据放在操作系统层面就显得相对高风险了。所以我们希望系统更及时地往磁盘写数据。

可以在`/etc/sysctl.conf`中加入下面两行，并执行"`sysctl -p`"

```
vm.dirty_background_ratio = 5
vm.dirty_ratio = 10
```

这是虚拟机的典型应用。不建议将它设置成0，毕竟有点后台IO可以提升一些程序的性能。

## 情景2：增加Cache

在一些场景中增加Cache是有好处的。例如，数据不重要丢了也没关系，而且有程序重复地读写一个文件。允许更多的cache，你可以更多地在内存上进行读写，提高速度。

```
vm.dirty_background_ratio = 50
vm.dirty_ratio = 80
```

有时候还会提高vm.dirty\_expire\_centisecs 这个参数的值，来允许脏数据更长时间地停留。

---

## 情景3：增减兼有

有时候系统需要应对突如其来的高峰数据，它可能会拖慢磁盘。（比如说，每个小时开始时进行的批量操作等）

这个时候需要容许更多的脏数据存到内存，让后台进程慢慢地通过异步方式将数据写到磁盘当中。

```
vm.dirty_background_ratio = 5  
vm.dirty_ratio = 80
```

这个时候，后台进行在脏数据达到5%时就开始异步清理，但在80%之前系统不会强制同步写磁盘。这样可以使IO变得更加平滑。

---

从/proc/vmstat, /proc/meminfo, /proc/sys/vm中可以获得更多资讯来作出调整。