

Re: Raid 5 performance

Source: <http://unix.derkeiler.com/Mailing-Lists/FreeBSD/performance/2004-02/0015.html>

From: Uwe Doering (gemini_at_geminix.org)

Date: 02/08/04

Date: Sun, 08 Feb 2004 00:36:51 +0100

To: freebsd-performance@freebsd.org

Allen Landsidel wrote:

> At 04:39 2/7/2004, Uwe Doering wrote:

>> Todd Lewis wrote:

>>> I am using FreeBSD 4.9, with a 3ware RAID 5

>>> 1 gig memory 2.8g p4

>>> Three questions.

>>> 1. FreeBSD has a 16k block size. The RAID card is set at 64k

>>> Block size(its sweet spot). My logic tells me that

>>> increasing the block size to 64k would increase disk

>>> read and write access. But, everything I read suggest

>>> going above 64k is dangerous. Are there any recommendations

>>> on performance a stability concerns when increasng the

>>> block size to 64k when using a RAID controller.

>>

>> A RAID controller normally has nothing to do with the file system's

>> block size. Are you sure that you're not mixing this up with the

>> stripe size? Which stripe size to use with a RAID controller depends

>> on your performance priorities.

>>

>> If there are a lot of concurrent disk operations a larger stripe size

>> is better because then a single disk operation tends to be limited to

>> only one disk drive, leaving the remaining drives free to perform

>> other and possibly unrelated disk operations at the same time. On the

>> other hand, if sequential i/o throughput is important a smaller stripe

>> size is better.

>

> I am compelled to step up here and say that this flies in the face not

> only of everything I have read, ever, about RAID -- but my own personal

> experiences as well on a variety of controllers and drives ranging from

> ATA highpoint/maxtor combos up to 4ch u160 ICP-Vortex/15krpm monsters.

>

> My experience has told me that for mostly sequential I/O, bigger is

> better, up to a point. 128KB stripes are much faster than 16KB stripes

> on every combination I've ever used when it comes to sequential I/O.

>

> Random I/O tends to prefer smaller stripe sizes because as you said,

> spreading things out over disks is better, and in random I/O the request

freebsd-performance: Re: Raid 5 performance

> *sizes tend to be much smaller.*

I didn't talk about random I/O on a single file but about parallel, unrelated disk operations (by different processes, for instance), each of which may very well be part of a sequential read or write. FreeBSD does clustering, so when an I/O operation actually takes place it can be much larger than the 16k block size. So in order to achieve a sufficiently high statistical likelihood that an I/O operation is limited to a single disk, leaving the remaining disks free for other I/O, you need a stripe size considerably larger than the file system's block size.

A file server with many parallel, unrelated accesses by NFS clients is a good example. There you certainly cannot generalize that files and therefore (clustered) I/O operations tend to be small. This generalization is only valid to some extent in case a file is a database where the I/O operations are mostly random and indeed tend to be small.

What all this means is that you need to carefully analyse the intended use and access pattern in order to pick the right stripe size, since there is no one-size-fits-all.

Uwe

--

Uwe Doering | EscapeBox - Managed On-Demand UNIX Servers
geminix@geminix.org | <http://www.escapebox.net>

freebsd-performance@freebsd.org mailing list
<http://lists.freebsd.org/mailman/listinfo/freebsd-performance>
To unsubscribe, send any mail to "freebsd-performance-unsubscribe@freebsd.org"