

Re: Filesystem snapshots dog slow

Source: <http://unix.derkeiler.com/Mailing-Lists/FreeBSD/hackers/2007-10/msg00155.html>

- *From:* Kostik Belousov <kostikbel@xxxxxxxxxx>
 - *Date:* Wed, 17 Oct 2007 13:14:00 +0300
-

On Wed, Oct 17, 2007 at 08:00:03PM +1000, Peter Jeremy wrote:

On 2007-Oct-16 06:54:11 -0500, Eric Anderson <anderson@xxxxxxxxxxxxx> wrote:

will give you a good understanding of what the issue is. Essentially, your disk is hammered making copies of all the cylinder groups, skipping those that are 'busy', and coming back to them later. On a 200Gb disk, you could have 1000 cylinder groups, each having to be locked, copied, unlocked, and then checked again for any subsequent changes. The stalls you see are when there are lock contentions, or disk IO issues. On a single disk (like your setup above), your snapshots will take forever since there is very little random IO performance available to you.

That said, there is a fair amount of scope available for improving both the creation and deletion performance.

Firstly, it's not clear to me that having more than a few hundred CGs has any real benefits. There was a massive gain in moving from (effectively) a single CG in pre-FFS to a few dozen CGs in FFS as it was first introduced. Modern disks are roughly 5 orders of magnitude larger and voice-coil actuators mean that seek times are almost independent of distance. CG sizes are currently limited by the requirement that the cylinder group (including cylinder group maps) must fit into a single FS block. Removing this restriction would allow CGs to be much larger.

Secondly, all the I/O during both snapshot creation and deletion is in FS-block size chunks. Increasing the I/O size would significantly increase the I/O performance. Whilst it doesn't make sense to read more than you need, there still appears to be plenty of scope to combine writes.

Between these two items, I would expect potential performance gains of at least 20:1.

Note that I'm not suggesting that either of these items is trivial.

This is, unfortunately, quite true. Allowing non-atomic updates of the

Re: Filesystem snapshots dog slow

cg block means a lot of complications in the softupdate code, IMHO.

Attachment: *pgpDIwCfUNaYy.pgp*

Description: PGP signature