Re: Performance problem (I/O)

Source: http://unix.derkeiler.com/Newsgroups/comp.unix.tru64/2004−03/0050.html

---

From: Bob Harris (*harris_at_zk3.dec.com*)
Date: 03/20/04

Date: Sat, 20 Mar 2004 00:01:40 GMT

In article <c3fijl$os9$1@jeeves.eng.abbnm.com>,
peter@abbnm.com (Peter da Silva) wrote:

> In article <harris−F51A96.21182918032004@cacnews.cac.cpqcorp.net>,
> Bob Harris <harris@zk3.dec.com> wrote:
> > When you restored the file system using tar, you placed all the files in
> > the exact order that you would then back them up. You positioned the
> > metadata for each file next to the file that would preceed it. As a
> > result you maximized your backup performance.
>
> AdvFS doesn't use an analog of the UFS cylinder group technique to keep
> metadata near the file data?

No.

> That would explain some of the performance problems we see in an app that
> writes a lot of small files.

Not having cylinder groups may not have anything to do with small file write performance. This could be due to the way the last 8K allocation of a small file is managed. For space efficency, files under 150K tend to have the last 8K of the file stored in a frag of from 1K to 7K in length. But while the is being written, a full 8K is allocated. When the file is closed, the size of the file is checked and if a 10% saving in space can be obtained by turning the last 8K into a frag, then a frag is allocated, the end of the flie copied to the frag, and then the original 8K is deallocated.

All of this results in additional disk I/O for small files when they are closed.

In the newer versions of Tru64 UNIX, there is chfsets option to disable this and make all files created from that point forward allocation storage in multiples of 8K. For older versions there is a global variable that can be patched in the kernel to disable frag'ing of a file.

In the 8 million file case, if the small files are evenly distributed between 1K and and any size that is 8K or greater, then turning off file

frag'ing would increase the storage usage for that file system by about
32 gigabytes. At one time I would have choaked on such a number.
Today, I have more storage than that on my laptop. I will not attempt
to place a value on this to you or your company as laptop storage is not
the same as SCSI, RAID, SAN storage which tends to come in smaller sizes
and cost more. But is still a much lower cost than when I was the
system manager for a VAX−11/780 :−) Times change.

Bob Harris