

Re: Bad performance when accessing a lot of small files

Re: Bad performance when accessing a lot of small files

Source: <http://unix.derkeiler.com/Mailing-Lists/FreeBSD/performance/2007-12/msg00076.html>

- *From:* "Alexandre Biancalana" <biancalana@xxxxxxxx>
 - *Date:* Fri, 21 Dec 2007 20:55:22 -0300
-

On 12/21/07, Alfred Perlstein <alfred@xxxxxxxx> wrote:

* Alexandre Biancalana <biancalana@xxxxxxxx> [071221 12:48] wrote:

On 12/21/07, Alfred Perlstein <alfred@xxxxxxxx> wrote:

Hi Alfred !

There is a lot of very good tuning advice in this thread,
however
one thing to note is that having ~1 million files in a directory
is not a very good thing to do on just about any filesystem.

I think I was not clear, I will try explain better.

This Backup Server has a /backup zfs filesystem of 4TB.

Each host that do backups to this server has a /backup/<hostname> and
/backup/<hostname>/YYYYMMDD zfs filesystems, the last contains the
backups for some day of that server.

My problem is with some hosts that have in your directory structure a
lot of small files, independent of the hierarchy.

Can you not tar these files together?

This is what I'm trying to do....

One trick that a lot of people do is hashing the directories
themselves
so that you use some kind of computation to break this huge

Re: Bad performance when accessing a lot of small files

Re: Bad performance when accessing a lot of small files

dir into
multiple smaller dirs.

I have the two cases, when you have a lot of files inside on directory without any directory organization/distribution but I also have problems with hosts that have files organized in a hierarchy like YYYY/MM/DD/<files> having no more that 200 files in the day directory level, but almost one million of files in total.

Just for info, I made the previous suggested tuning (raise dirhash, maxvnodes) but this improve nothing.

Thanks for your hint!

What application are you scanning these files with? I know I had issues with rsync in particular where I had to have it rsync smaller pieces of a collection for it to work nicely instead of going for the whole heirarchy.

tar

I run tar in the /backup/<hostname>/YYYYMMDD writing to LTO3 tape drive, the problem is that when origin directory contains a lot of small files the process is *much* more slow.... this is my question since the thread start.

freebsd-performance@xxxxxxxxxxx mailing list

<http://lists.freebsd.org/mailman/listinfo/freebsd-performance>

To unsubscribe, send any mail to "freebsd-performance-unsubscribe@xxxxxxxxxxx"