

Re: copying large files between filesystems

Source: <http://linux.derkeiler.com/Mailing-Lists/Fedora/2004-06/8098.html>

From: Andrew Scott (ascott_at_tathata.org)

Date: 06/30/04

Date: Wed, 30 Jun 2004 13:55:25 -0700

To: For users of Fedora Core releases <fedora-list@redhat.com>

Phil Schaffner wrote:

[snip]

> *On Tue, 2004-06-29 at 18:43 -0700, Andrew Scott wrote:*

[snip]

> *IIRC older resiserfs (version < 3.6) had a 2GB limit, so it depends.*
> *Seems like it should have given an error. The limit on ext3 file size*
> *depends on the kernel, so again it depends on the details of your*
> *system. Particular tools may also have built-in filesize limits – have*
> *been some discussions about getting large files such as DVD images with*
> *tools such as wget. Archive tools may have the same issues.*
>

debugreiserfs reports that it is a 3.6 filesystem.

>

>>*Immediately after I made*

>>*the original bzip2 archive I ran md5sum on the resulting file, and kept*
>>*a copy of that number.*

>

>

> *Not quite clear on the archive issue – bzip2 is a compression method,*
> *not an archive tool. What kind of archive is it? Can you get a*
> *directory of the original archive in place? (e.g. For a tar archive,*
> *could you do "tar jtvf archive_name.tar.bz"?*

>

it's a tarball tha has been bziped

>

>>*After I copy the file to another filesystem, I*

>>*can not get the same md5sum. I've tried cp, rsync, scp, and dd and they*
>>*each seem to come up with different md5sums. The byte count is the same*
>>*however, but the md5sum is always different, so something however*
>>*minimal is changing and henceforth throwing off the whole bzip2 archive*
>>*and rendering it unextractable.*

>

>

Fedora: Re: copying large files between filesystems

> *If you can read the directory in place, you may be able to extract to*
> *the new target location (at least up to the 2GB limit if not beyond)*
> *without moving the archive.*
>
> *(e.g. "tar jxvf archive_name.tar.bz -C /new/filesystem/location")*
>

It actually looks now that the problem is not filesystem but that the drive is throwing errors intermittantly. :-/

>
>>*I'm totally stressed because this is a*
>>*back up of my homedir from before a reinstall. :-/ And I'm worried*
>>*that the original filesystem that I wrote the backup to (the reiserfs*
>>*filesystem) was so old that it didn't recognize files greater than 2*
>>*Gigs, silently.*
>>
>>*If I md5sum the file, still in it's place, it's still good.*
>
>
> *But if it's already effectively truncated at 2GB it may not actually be*
> *good.*
>

I think if the reiserfs filesystem had a 2G limit then the original copy to that drive would have failed, but it didn't. I think the problem is solely the result of bad I/O from the drive now and my problem has become that much more complicated.

>
>>*Any ideas why I can't get matching md5sums from a file copy between*
>>*filesystems of different type? Any ideas how to recover from this?*
>>
>>*Thanks in advance, anyone, with thoughts,*
>
>
> *Provide more details and you may be able to get better help.*
>

I'm currently on Fedora core 2 with a 2.6.6 kernel. I'll have to try unarchiving to another filesystem. I'm runnig badblocks on the drive right now. I've freed up enough space on the drive to uncompress in place, but that failed with I/O erros. Then I tried the bzip2recover program with absolutely horrible results as it creates over 2000 9K bzip2 files representing each 9k block in the archive on the drive which was so taxing on the drive that it caused it cough and spit.

Any ideas how to do a really slow read from a drive that might prove more accurate (less taxing) on the hardware? I've tried dd and am now thinking about resorting to running strings on the device and piping it to another filesystem, but that will probably still have errors in the

Fedora: Re: copying large files between filesystems

resulting file.

I emailed the guys at Namesys (reiserfs headquarters in Oakland, CA). They have a standing offer of "Ask any questions for \$25". I sent them \$25 and asked them a question. Hans Reiser got back to me as well as another employee, both with good suggestions. They suspected the hardware immediately. They made one really keen suggestion: if the bit count is identical on the original as the copy (when copied to another filesystem), but the md5sums are different, then try and run bindiff on the two files and use a binary editor to toggle the differing bits, with the goal of a correct md5sum match. I imagine this will be the last thing I try before sending the disk off for disk recovery.

Anyway, thanks a lot for your time and thoughts. What a pain in the ass.

–Andrew

> *How To Ask Questions The Smart Way*
> <http://www.catb.org/~esr/faqs/smart-questions.html>
>
> *Phil*
>
>
>
>

--

fedora-list mailing list
fedora-list@redhat.com

To unsubscribe: <http://www.redhat.com/mailman/listinfo/fedora-list>