

Re: 2.6.19 file content corruption on ext3

Source: <http://linux.derkeiler.com/Mailing-Lists/Kernel/2006-12/msg07755.html>

- *From:* Andrew Morton <akpm@xxxxxxxx>
 - *Date:* Thu, 28 Dec 2006 17:59:54 -0800
-

On Thu, 28 Dec 2006 17:38:38 -0800 (PST)
Linus Torvalds <torvalds@xxxxxxxx> wrote:

in
the hope that somebody else is working on this corruption issue and is
interested..

What corruption issue? ;)

I'm finding that the corruption happens trivially with your test app, but
apparently doesn't happen at all with ext2 or ext3, data=writeback. Maybe
it will happen with increased rarity, but the difference is quite stark.

Removing the

```
err = walk_page_buffers(handle, page_bufs, 0, PAGE_CACHE_SIZE,  
NULL, journal_dirty_data_fn);
```

from ext3_ordered_writepage() fixes things up.

The things which journal_submit_data_buffers() does after dropping all the
locks are ... disturbing – I don't think we have sufficient tests in there
to ensure that the buffer is still where we think it is after we retake
locks (they're slippery little buggers). But that wouldn't explain it
anyway.

It's inefficient that journal_dirty_data() will put these locked, clean
buffers onto BJ_SyncData instead of BJ_Locked, but
journal_submit_data_buffers() seems to dtrt with them.

So no theory yet. Maybe ext3 is just altering timing. But the difference
is really large..

Disabling all the WB_SYNC_NONE stuff and making everything go synchronous

Re: 2.6.19 file content corruption on ext3

everywhere has no effect. Disabling bdi_write_congested() has no effect.

—

To unsubscribe from this list: send the line "unsubscribe linux-kernel" in the body of a message to majordomo@xxxxxxxxxxxxxxxx

More majordomo info at <http://vger.kernel.org/majordomo-info.html>

Please read the FAQ at <http://www.tux.org/lkml/>