

Re: [00/17] Large Blocksize Support V3

Source: <http://linux.derkeiler.com/Mailing-Lists/Kernel/2007-04/msg09688.html>

- *From:* David Chinner <dgc@xxxxxxx>
 - *Date:* Fri, 27 Apr 2007 00:18:26 +1000
-

On Thu, Apr 26, 2007 at 01:37:42PM +0100, Andy Whitcroft wrote:

Nick Piggin wrote:

Christoph Lameter wrote:

On Thu, 26 Apr 2007, Nick Piggin wrote:

What are the exact
requirement you are trying
to address?

Block size > page cache size.

But what do you mean with it? A block is no longer a
contiguous
section of memory. So you have redefined the term.

I don't understand what you mean at all. A block has always been a
contiguous area of disk.

Lets take Nick's definition of block being a disk based unit for the
moment. That does not change the key contention here, that even with
hardware specifically designed to handle 4k pages that hardware handles
larger contiguous areas more efficiently. David Chinner gives us
figures showing major overall throughput improvements from (I assume)
shorter scatter gather lists and better tag utilisation.

I haven't actually provided any figures – it's knowledge passed down
from those that know more about it that I do.

If you want figures about the impact of large I/Os, then we should
not be looking at the HBAs but at the impact on RAID controller
throughput (this I do have numbers on ;). It is not uncommon to

Re: [00/17] Large Blocksize Support V3

see 2MB I/Os give twice the throughput of 512K I/Os to a single RAID controller – larger than 512k can only be achieved on systems with a page size larger than 4k.....

I am loath to say we can just blame the hardware vendors for poor design.

Never did – I'm pointing out that linux can't use all the capabilities they have.

Cheers,

Dave.

—

Dave Chinner
Principal Engineer
SGI Australian Software Group

—

To unsubscribe from this list: send the line "unsubscribe linux-kernel" in the body of a message to majordomo@xxxxxxxxxxxxxxxxx
More majordomo info at <http://vger.kernel.org/majordomo-info.html>
Please read the FAQ at <http://www.tux.org/lkml/>