

Re: [PATCH 0/13] maps: pagemap, kpagemap, and related cleanups

Re: [PATCH 0/13] maps: pagemap, kpagemap, and related cleanups

Source: <http://linux.derkeiler.com/Mailing-Lists/Kernel/2007-04/msg04867.html>

- *From:* Matt Mackall <mpm@xxxxxxxxxxxx>
 - *Date:* Fri, 13 Apr 2007 12:24:51 -0500
-

On Fri, Apr 13, 2007 at 10:03:56AM -0700, Andrew Morton wrote:

On Fri, 13 Apr 2007 11:24:36 -0500 Matt Mackall <mpm@xxxxxxxxxxxx> wrote:

It **will** be viable. If the application wants to know if a page is dirty, it looks up "PG_dirty" in /proc/pg_foo-to-bitnumber and uses PG_dirty's numerical offset when inspecting fields in /proc/kpagemap. If correctly designed, such a monitoring application will be able to report upon page flags which we haven't even thought up yet.

We can probably fit this in the existing (variable-sized) header.

hm, OK..

I wonder what they are needed for.

Poking deeply into the kernel to provide information about kernel state.

There are real-world needs for this, and the people who develop tools to process this information will have decent kernel understanding and will know that the file's contents may alter across kernel versions. It sure beats poking around in /dev/kmem.

I doubt if there's a sensible way in which we can prettify this interface

Re: [PATCH 0/13] maps: pagemap, kpagemap, and related cleanups

without losing information. But we should aim to make it as robust as possible against future kernel changes, of course.

And we should satisfy ourselves that all the required information has been made available. The fact that it will satisfy the Oracle requirement is encouraging.

Matt, these changes make the new field in /proc/pid/smmaps redundant, don't they?

Which new field?

Referenced:

From /proc/kpagemap + /proc/*/pagemap, you can basically synthesize any statistic you want, including all the existing ones. For some data, /proc/pid/smmaps (or /proc/meminfo) will be considerably more efficient.

You'd need to poke clear_refs beforehand to make the referenced bits useful.

Actually, we also need to run around the ptes and collect the pte-referenced bits too. I don't think your code copes with any of that?

No, and it probably should. Perhaps dirty as well, though I've kind of lost the plot on how that works lately.

But in general, most of the statistics in smmaps are basically useless for shared mappings, just like RSS. Problem is, we really don't know what statistics we want yet, or even if it can be distilled down to simple numbers anyway.

yup. But that's the whole point, really: don't prejudge what info userspace is trying to collect.

Right.

--

Mathematics is the supreme nostalgia of our time.

Re: [PATCH 0/13] maps: pagemap, kpagemap, and related cleanups

Re: [PATCH 0/13] maps: pagemap, kpagemap, and related cleanups

—

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

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

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