
Subject: Re: [PATCH] fix bad behavior in use_hierarchy file
Posted by [Glauber Costa](#) on Tue, 26 Jun 2012 11:12:07 GMT
[View Forum Message](#) <> [Reply to Message](#)

On 06/26/2012 03:10 PM, Michal Hocko wrote:
> On Tue 26-06-12 14:31:51, Glauber Costa wrote:
>> On 06/26/2012 11:56 AM, Michal Hocko wrote:
>>> [Adding Ying to CC - they are using hierarchies AFAIU in their workloads]
>>>
>>> On Mon 25-06-12 13:49:08, Tejun Heo wrote:
>>> [...]
>>>> A bit of delta but is there any chance we can either deprecate
>>>> .use_hierarchy or at least make it global toggle instead of subtree
>>>> thing?
>>>
>>> So what you are proposing is to have all subtrees of the root either
>>> hierarchical or not, right?
>>>
>>>> This seems needlessly complicated. :(
>>>
>>> Toggle wouldn't help much I am afraid. We would still have to
>>> distinguish (non)hierarchical cases. And I am not sure we can make
>>> everything hierarchical easily.
>>> Most users (from my experience) ignored use_hierarchy for some reasons
>>> and the end results might be really unexpected for them if they used
>>> deeper subtrees (which might be needed due to combination with other
>>> controller(s)).
>>>
>> Do we have any idea about who those users are, and how is their
>> setup commonly done?
>
> Well, most of them use memory controller with combination of other
> controller - usually cpuset or cpu - and memcg is used to cap the amount
> of memory for each respective group. As I said most of those users
> were not aware of use_hierarchy at all.
>
>> We can propose work arounds here, but not without first knowing work
>> arounds to what =p
>
> No, please no workarounds. It will be even bigger mess.
> Maybe a global switch is the first step in the right direction (on by
> default). If somebody encounters any issue we can say it can be turned
> off (something like one time switch) or advise on how to fix their
> layout to fit hierarchy better. We can put WARN_ON_ONCE when the knob is
> set to 0 in the second stage and finally remove the whole knob.
>

Sorry for the wording. I didn't mean work around in the sense of a

kludge. I meant it as actually proposing solutions to the problem that would disrupt people as little as we can.

Well, instead of a global switch, a much easier thing would be to set it to 1 by default. It would actually work as a global switch, because we always inherit the parent's value.

You can set the root to 0 before you add other groups, but that generates a warning, as you suggested.

But after it was first set to 0, he would be free to keep using mixed configurations if needed - this way we're likely to find out if there are actually users of that around.
