Subject: Re: containers (was Re: -mm merge plans for 2.6.23)
Posted by Srivatsa Vaddagiri on Wed, 11 Jul 2007 10:03:23 GMT
View Forum Message <> Reply to Message

On Wed, Jul 11, 2007 at 02:23:52AM -0700, Paul Jackson wrote:
> Ingo wrote:
> > another option would be to trivially hook up CONFIG_FAIR_GROUP_SCHED
> > with cpusets, ...
>
> ah ... you triggered my procmail filter for 'cpuset' ... ;).

:-)

> What would it mean to hook up CFS with cpusets?

CFS is the new cpu scheduler in Linus's tree (http://lwn.net/Articles/241085/).
It has some group scheduling capabilities added i.e the core scheduler
now recognizes the concept of a task-group and providing fair cpu time
to each task-group (in addition to providing fair time to each task in a
group).

The core scheduler however is not concerned with how task groups are formed
and/or how tasks migrate between groups. Thats where a patch like Paul Menage's
container infrastructure comes in hand - to provide a user-interface for
managing task-groups (create/delete task groups, migrate task from one
group to another etc). Whatever the chosen user-interface is, cpu
scheduler needs to know about such task-group creation/destruction,
migration of tasks across groups etc.

Unfortunately, the group-scheduler bits will be ready in 2.6.23 while
Paul Menage's container patches aren't ready for 2.6.23 yet.

So Ingo was proposing we use cpuset as that user interface to manage
task-groups. This will be only for 2.6.23. In 2.6.24, when hopefully Paul
Menage's container patches will be ready and will be merged, the group
cpu scheduler will stop using cpuset as that interface and use the
container infrastructure instead.

If you recall, I have attempted to use cpuset for such an interface in
the past (metered cpusets - see [1]). It brings in some semantic changes
for cpusets, most notably:

- metered cpusets cannot have grand-children
- all cpusets under a metered cpuset need to share the same set
  of cpus.

Is it fine if I introduce these semantic changes, only for 2.6.23 and
only when CONFIG_FAIR_GROUP_SCHED is enabled? This will let the group

cpu scheduler to receive some amount of testing.

The other alternative is to hook up group scheduler with user-id's
(again only for 2.6.23).

> I've a pretty
> good idea what a cpuset is, but don't know what kind of purpose
> you have in mind for such a hook. Could you say a few words to
> that? Thanks.

Reference:

1. http://marc.info/?l=linux-kernel&m=115946525811848&w=2

--
Regards,
vatsa

_____

Containers mailing list
Containers@lists.linux-foundation.org
https://lists.linux-foundation.org/mailman/listinfo/containers