Subject: Re: [patch00/05]: Containers(V2)- Introduction
Posted by Rohit Seth on Wed, 20 Sep 2006 18:07:58 GMT
View Forum Message <> Reply to Message

On Wed, 2006-09-20 at 10:38 -0700, Christoph Lameter wrote:
> On Wed, 20 Sep 2006, Rohit Seth wrote:
>
> > cpusets provides cpu and memory NODES binding to tasks. And I think it
> > works great for NUMA machines where you have different nodes with its
> > own set of CPUs and memory. The number of those nodes on a commodity HW
> > is still 1. And they can have 8-16 CPUs and in access of 100G of
> > memory. You may start using fake nodes (untested territory) to
>
> See linux-mm. We just went through a series of tests and functionality
> wise it worked just fine.
>

I thought the fake NUMA support still does not work on x86_64 baseline
kernel. Though Paul and Andrew have patches to make it work.

> > translate a single node machine into N different nodes. But am not sure
> > if this number of nodes can change dynamically on the running machine or
> > a reboot is required to change the number of nodes.
>
> This is commonly discussed under the subject of memory hotplug.
>

So now we depend on getting memory hot-plug to work for faking up these
nodes ...for the memory that is already present in the system. It just
does not sound logical.

> > Though when you want to have in access of 100 containers then the cpuset
> > function starts popping up on the oprofile chart very aggressively. And
> > this is the cost that shouldn't have to be paid (particularly) for a
> > single node machine.
>
> Yes this is a new way of using cpusets but it works and we could fix the
> scalability issues rather than adding new subsystems.
>

I think when you have 100's of zones then cost of allocating a page will
include checking cpuset validation and different zone list traversals
and checks...unless there is major surgery.

> > And what happens when you want to have cpuset with memory that needs to
> > be even further fine grained than each node.
>
> New node?

>

Am not clear how is this possible. Could you or Paul please explain.

> > Containers also provide a mechanism to move files to containers. Any
> > further references to this file come from the same container rather than
> > the container which is bringing in a new page.
>
> Hmmmm... Thats is interesting.
>
> > In future there will be more handlers like CPU and disk that can be
> > easily embeded into this container infrastructure.
>
> I think we should have one container mechanism instead of multiple. Maybe
> merge the two? The cpuset functionality is well established and working
> right.
>

I agree that we will need one container subsystem in the long run.
Something that can easily adapt to different configurations.

-rohit
