Subject: Re: [ckrm-tech] [patch00/05]: Containers(V2)- Introduction
Posted by Chandra Seetharaman on Wed, 20 Sep 2006 18:54:56 GMT
View Forum Message <> Reply to Message

On Wed, 2006-09-20 at 11:43 -0700, Paul Menage wrote:
> On 9/20/06, Chandra Seetharaman <sekharan@us.ibm.com> wrote:
> > > We already have such a functionality in the kernel its called a cpuset. A
> >
> > Christoph,
> >
> > There had been multiple discussions in the past (as recent as Aug 18,
> > 2006), where we (Paul and CKRM/RG folks) have concluded that cpuset and
> > resource management are orthogonal features.
> >
> > cpuset provides "resource isolation", and what we, the resource
> > management guys want is work-conserving resource control.
>
> CPUset provides two things:
>
> - a generic process container abstraction
>
> - "resource controllers" for CPU masks and memory nodes.
>
> Rather than adding a new process container abstraction, wouldn't it
> make more sense to change cpuset to make it more extensible (more
> separation between resource controllers), possibly rename it to
> "containers", and let the various resource controllers fight it out
> (e.g. zone/node-based memory controller vs multiple LRU controller,
> CPU masks vs a properly QoS-based CPU scheduler, etc)
>
> Or more specifically, what would need to be added to cpusets to make
> it possible to bolt the CKRM/RG resource controllers on to it?

Paul,

We had this discussion more than 18 months back and concluded that it is
not the right thing to do. Here is the link to the thread:

http://marc.theaimsgroup.com/?t=109173653100001&r=1& w=2

chandra
>
> Paul
>
> ---------- ----------
> Take Surveys. Earn Cash. Influence the Future of IT
> Join SourceForge.net's Techsay panel and you'll get the chance to share your
> opinions on IT & business topics through brief surveys -- and earn cash

> http://www.techsay.com/default.php?page=join.php&p=sourc eforge&CID=DEVDEV
> _____
> ckrm-tech mailing list
> https://lists.sourceforge.net/lists/listinfo/ckrm-tech
--

---------- ----------
Chandra Seetharaman | Be careful what you choose....
- sekharan@us.ibm.com | .....you may get it.
---------- ----------

Subject: Re: [ckrm-tech] [patch00/05]: Containers(V2)- Introduction
Posted by Paul Menage on Wed, 20 Sep 2006 19:25:15 GMT
View Forum Message <> Reply to Message

On 9/20/06, Chandra Seetharaman <sekharan@us.ibm.com> wrote:
>
> We had this discussion more than 18 months back and concluded that it is
> not the right thing to do. Here is the link to the thread:

Even if the resource control portions aren't totally compatible,
having two separate process container abstractions in the kernel is
sub-optimal, both in terms of efficiency and userspace management. How
about splitting out the container portions of cpuset from the actual
resource control, so that CKRM/RG can hang off of it too? Creation of
a cpuset or a resource group would be driven by creation of a
container; at fork time, a task inherits its parent's container, and
hence its cpuset and/or resource groups.

At its most crude, this could be something like:

```
struct container {
#ifdef CONFIG_CPUSETS
  struct cpuset cs;
#endif
#ifdef CONFIG_RES_GROUPS
  struct resource_group rg;
#endif
};
```

but at least it would be sharing some of the abstractions.

Paul

Subject: Re: [ckrm-tech] [patch00/05]: Containers(V2)- Introduction

Posted by Christoph Lameter on Wed, 20 Sep 2006 19:55:25 GMT
View Forum Message <> Reply to Message

On Wed, 20 Sep 2006, Chandra Seetharaman wrote:

> We had this discussion more than 18 months back and concluded that it is
> not the right thing to do. Here is the link to the thread:

Recent discussions on linux-mm sounded very different. I also brought this up at the VM summit. Could you have a look at cpusets and the discussion on linux-mm and then think how this could be done in a less VM invasive way?

Subject: Re: [ckrm-tech] [patch00/05]: Containers(V2)- Introduction
Posted by Paul Jackson on Wed, 20 Sep 2006 20:27:34 GMT
View Forum Message <> Reply to Message

Chandra wrote:
> We had this discussion more than 18 months back and concluded that it is
> not the right thing to do. Here is the link to the thread:

Because it is easy enough to carve memory up into nice little nameable chunks, it might be the case that we can manage the percentage of memory used by the expedient of something like cpusets and fake nodes.

Indeed, that seems to be doable, based on this latest work of Andrew and others (David, some_bright_spark@jp, Magnus, ...). There are still a bunch of wrinkles that remain to be ironed out.

For other resources, such as CPU cycles and network bandwidth, unless another bright spark comes up with an insight, I don't see how to express the "percentage used" semantics provided by something such as CKRM, using anything resembling cpusets.

... Can one imagine having the scheduler subdivide each second of time available on a CPU into several fake-CPUs, each one of which speaks for one of those sub-second fake-CPU slices? Sounds too weird to me, and a bit too rigid to be a servicable CKRM substitute.

--

I won't rest till it's the best ...
Programmer, Linux Scalability
Paul Jackson <pj@sgi.com> 1.925.600.0401

Subject: Re: [ckrm-tech] [patch00/05]: Containers(V2)- Introduction

Posted by Srivatsa Vaddagiri on Thu, 21 Sep 2006 17:02:26 GMT
View Forum Message <> Reply to Message

On Wed, Sep 20, 2006 at 01:27:34PM -0700, Paul Jackson wrote:
> For other resources, such as CPU cycles and network bandwidth, unless
> another bright spark comes up with an insight, I don't see how to
> express the "percentage used" semantics provided by something such
> as CKRM, using anything resembling cpusets.

How abt metered cpusets? Each child cpuset of a metered cpuset
represents a fraction of CPU time alloted to the tasks of the child
cpuset.

> ... Can one imagine having the scheduler subdivide each second of
> time available on a CPU into several fake-CPUs, each one of which
> speaks for one of those sub-second fake-CPU slices? Sounds too
> weird to me, and a bit too rigid to be a servicable CKRM substitute.

--
Regards,
vatsa

Subject: Re: [ckrm-tech] [patch00/05]: Containers(V2)- Introduction
Posted by Paul Jackson on Thu, 21 Sep 2006 19:29:00 GMT
View Forum Message <> Reply to Message

Vatsa wrote:
> How abt metered cpusets? Each child cpuset of a metered cpuset
> represents a fraction of CPU time alloted to the tasks of the child
> cpuset.

Ah yes - they might work. Sorry I didn't think of your
meter_cpu controller patch with its cpuset interface
when I wrote the above.

--

I won't rest till it's the best ...
Programmer, Linux Scalability
Paul Jackson <pj@sgi.com> 1.925.600.0401
