Subject: Re: [ckrm-tech] [PATCH 4/7] UBC: syscalls (user interface)
Posted by Chandra Seetharaman on Fri, 18 Aug 2006 18:27:34 GMT
View Forum Message <> Reply to Message

On Fri, 2006-08-18 at 11:17 -0700, Chandra Seetharaman wrote:
> On Fri, 2006-08-18 at 09:42 -0700, Andrew Morton wrote:
> > On Fri, 18 Aug 2006 07:45:48 -0700
> > Dave Hansen <haveblue@us.ibm.com> wrote:
> >
> > > On Fri, 2006-08-18 at 12:08 +0400, Andrey Savochkin wrote:
> > > >
> > > > A) Have separate memory management for each container,
> > > >   with separate buddy allocator, lru lists, page replacement mechanism.
> > > >   That implies a considerable overhead, and the main challenge there
> > > >   is sharing of pages between these separate memory managers.
> > >
> > > Hold on here for just a sec...
> > >
> > > It is quite possible to do memory management aimed at one container
> > > while that container's memory still participates in the main VM.
> > >
> > > There is overhead here, as the LRU scanning mechanisms get less
> > > efficient, but I'd rather pay a penalty at LRU scanning time than divide
> > > up the VM, or coarsely start failing allocations.
> > >
> >
> > I have this mad idea that you can divide a 128GB machine up into 256 fake
> > NUMA nodes, then you use each "node" as a 512MB unit of memory allocation.
> > So that 4.5GB job would be placed within an exclusive cpuset which has nine
> > "mems" (what are these called?) and voila: the job has a hard 4.5GB limit,
> > no kernel changes needed.
>
> In this model memory and container are tightly coupled, hence memory
> might be unused/wasted in one container/resource group", while a
> different group is hitting its limit too often.
>
> In order to minimize this effect, resource controllers should be
> providing both minimum and maximum amount of resources available for a
> resource group.

Forgot to mention... There is a set of patches submitted by KUROSAWA Takahiro that implements this by creating pseudo zones (It has the same limitation though). http://marc.theaimsgroup.com/?l=ckrm-tech&m=113867467006531&w=2

>
> >
> > Unfortunately this is not testable because numa=fake=256 doesn't come even

> > vaguely close to working. Am trying to get that fixed.
> >
> > ---------- ----------
> > Using Tomcat but need to do more? Need to support web services, security?
> > Get stuff done quickly with pre-integrated technology to make your job easier
> > Download IBM WebSphere Application Server v.1.0.1 based on Apache Geronimo
> > http://sel.as-us.falkag.net/sel?cmd=lnk&kid=120709&b id=263057&dat=121642
> > _____
> > 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.
---------- ----------