Subject: Re: [ckrm-tech] [PATCH] BC: resource beancounters (v4) (added user memory)
Posted by Pavel Emelianov on Tue, 12 Sep 2006 10:40:03 GMT
View Forum Message <> Reply to Message

Balbir Singh wrote:
> Pavel Emelianov wrote:
>> Balbir Singh wrote:
>>> Pavel Emelianov wrote:
>>>> Balbir Singh wrote:
>>>>> Dave Hansen wrote:
>>>>>> On Fri, 2006-09-08 at 11:33 +0400, Pavel Emelianov wrote:
>>>>>>> I'm afraid we have different understandings of what a
>>>>>>> "guarantee" is.
>>>>>> It appears so.
>>>>>>
>>>>>>> Don't we?
>>>>>>> Guarantee may be one of
>>>>>>>
>>>>>>>   1. container will be able to touch that number of pages
>>>>>>>   2. container will be able to sys_mmap() that number of pages
>>>>>>>   3. container will not be killed unless it touches that number of
>>>>>>> pages
>>>>>> A "death sentence" guarantee? I like it. :)
>>>>>>
>>>>>>>   4. anything else
>>>>>>>
>>>>>>> Let's decide what kind of a guarantee we want.
>>>>> I think of guarantees w.r.t resources as the lower limit on the
>>>>> resource.
>>>>> Guarantees and limits can be thought of as the range (guarantee,
>>>>> limit]
>>>>> for the usage of the resource.
>>>>>
>>>>>> I think of it as: "I will be allowed to use this many total
>>>>>> pages, and
>>>>>> they are guaranteed not to fail." (1), I think. The sum of all of
>>>>>> the
>>>>>> system's guarantees must be less than or equal to the amount of free
>>>>>> memory on the machine.
>>>>> Yes, totally agree.
>>>> Such a guarantee is really a limit and this limit is even harder than
>>>> BC's one :)
>>>>
>>>> E.g. I have a node with 1Gb of ram and 10 containers with 100Mb
>>>> guarantee each.
>>>> I want to start one more. What shall I do not to break guarantees?
>>> Don't start the new container or change the guarantees of the existing

>>> ones
>>> to accommodate this one :) The QoS design (done by the administrator)
>>> should
>>> take care of such use-cases. It would be perfectly ok to have a
>>> container
>>> that does not care about guarantees to set their guarantee to 0 and set
>>> their limit to the desired value. As Chandra has been stating we
>>> need two
>>> parameters (guarantee, limit), either can be optional, but not both.
>> If I set up 9 groups to have 100Mb limit then I have 100Mb assured (on
>> 1Gb node)
>> for the 10th one exactly. And I do not have to set up any guarantee as
>> it won't affect
>> anything. So what a guarantee parameter is needed for?
>
> This use case works well for providing guarantee to one container.
> What if
> I want guarantees of 100Mb and 200Mb for two containers? How do I setup
> the system using limits?
You may set any value from 100 up to 800 Mb for the first one and
200-900Mb for
the second. In case of no other groups first will receive its 100Mb for
sure and
so does the second. If there are other groups - their guarantees should
be concerned.
>
> Even I restrict everyone else to 700Mb. With this I cannot be sure that
> the remaining 300Mb will be distributed as 100Mb and 200Mb.
There's no "everyone else" here - we're talking about a "static" case.
When new group arrives we need to recalculate guarantees as you said.
And here's my next question - what to do if the new guarantee would become
lower that current amount of unreclaimable memory in BC?