Subject: Disabling all limits
Posted by divB on Wed, 08 Apr 2009 23:11:39 GMT
View Forum Message <> Reply to Message

Hello,

I'm migrating from Linux VServer to OpenVZ. With VServer I did not setup any limits, so the hardware was shared among all virtual servers in a "first come, first serve" manner.

Now I sometimes get the error:

-bash: fork: Cannot allocate memory

I think it's because of the resource limit.

Is there a way to switch off all limits for all containers so that they behave the same as with VServer?

I would only do this temporarily in order to get the servers quickly up'n'running and would restore and adapt the limits afterwards...

I just do not want to have errors in my migration process which are just caused by limits (I have enough troubles with the migration)

Thank you,
divB

Subject: Re: Disabling all limits
Posted by maratrus on Thu, 09 Apr 2009 06:43:54 GMT
View Forum Message <> Reply to Message

Hello,

Quote:
Now I sometimes get the error:

/proc/user_beancounters is the file you should look at.
Increasing counters in failcnt column (the last column) point you to resource you have to increase.

Quote:
Is there a way to switch off all limits for all containers so that they behave the same as with VServer?

I don't know how to achieve this in general situation but you can always adjust per VE configuration file with a quite big values.
You can also create the single config-sample file and then create VEs which will be based on that single file (vzctl create $VEID --config)
http://download.openvz.org/doc/OpenVZ-Users-Guide.pdf (page 33)

Subject: Re: Disabling all limits
Posted by divB on Thu, 09 Apr 2009 22:21:12 GMT
View Forum Message <> Reply to Message

Thank you very much, I decided do not remove the limits but rather adapt the needed values manually.

But, please help me a little bit: What is the best (i.e. empiric) way to adopt the limits according to the "failcnt" value? It seems that just adding this value to the soft/hardlimit isn't enough...

What I do: I have written a small awk-script which monitors the sum of the failcnt-column in /proc/user_beancounters and if this value is >= 0 (i.e. some hosts have insufficient resources) send the output via cron.

I have a VE with the following content:

privvmpages 68769 90280 67536 71632 97

(and PRIVVMPAGES="87536:91632" in the config).

This means that this value is just ~100 too low; what to I do? vzctl set 201 --privvmpages 87536:91632 --save and restarting the VE. As you can see, I add 10000 (!) although only 97 are needed. But some time after restarting: same result. I adjusted the limit a few times now but failcnt becomes always nonzero for this specific limit after some time. So, what I am doing wrong?

There are other values in other VEs where I have the same problem, e.g.:

kmemsize 5122700 11059046 11055923 11377049 1434

And, one other thing. If I stop a VE, the entries in /proc/user_beancounters are not removed (only after some time). Why? This is especially problematic when I want to correct the limits, do "vzctl 201 restart" and check if the problems have gone.

I hope you can follow me

Regards,
divB

Subject: Re: Disabling all limits
Posted by maratrus on Fri, 10 Apr 2009 06:50:27 GMT
View Forum Message <> Reply to Message

Hi,

Quote:
What I do: I have written a small awk-script which monitors the sum of the failcnt-column in /proc/user_beancounters and if this value is >= 0

This approach is not correct. You should monitor if faicnt are increasing not if they are just >0.

Quote:
... and restarting the VE.

after you've adjusted ubc-parameter you don't have to restart VE.

Quote:
This means that this value is just ~100 too low; what to I do?

vzctl set 201 --privvmpages 87536:91632 --save

and restarting the VE. As you can see, I add 10000 (!) although only 97 are needed. But some time after restarting: same result. I adjusted the limit a few times now but failcnt becomes always nonzero for this specific limit after some time. So, what I am doing wrong?

Sorry, I didn't catch you at all. What are 100 and 10000? What kind of parameters do you measure?

Quote:
And, one other thing. If I stop a VE, the entries in /proc/user_beancounters are not removed (only after some time)

This might be a valid behavior. Here is a nice article.
http://wiki.openvz.org/UBC_failcnt_reset

Subject: Re: Disabling all limits
Posted by divB on Fri, 10 Apr 2009 09:57:52 GMT
View Forum Message <> Reply to Message

maratrus wrote on Fri, 10 April 2009 02:50Hi,
Quote:
What I do: I have written a small awk-script which monitors the sum of the failcnt-column in /proc/user_beancounters and if this value is >= 0

This approach is not correct. You should monitor if faicnt are increasing not if they are just >0.

Quote:
... and restarting the VE.

after you've adjusted ubc-parameter you don't have to restart VE.

I think we have a misunderstanding here. In my case, the goal is to find out how much ressources are needed by my servers. So I just want to do a short "learning phase" (i.e. 1 week) and observe the "failcnt" field. If it is > 0 then my VE need more ressources. I adjust the parameters, restart and hope that the VEs work now normal.

There are at least two use cases for OpenVZ:

1.) Maybe the most popular: vServer hosting. In this case it's good to strictly apply ressource barriers.

2.) Virtualize private servers. This is my case. In normal conditions there should not be any malicous software; it's better to be sure that all my own services run reliably. So in this case I want each VE to give as much ressources as they need.

Quote:
This means that this value is just ~100 too low; what to I do?

vzctl set 201 --privvmpages 87536:91632 --save

and restarting the VE. As you can see, I add 10000 (!) although only 97 are needed. But some time after restarting: same result. I adjusted the limit a few times now but failcnt becomes always nonzero for this specific limit after some time. So, what I am doing wrong?

Sorry, I didn't catch you at all. What are 100 and 10000? What kind of parameters do you measure?
[/quote]

This is my key question: What to do if the ressources are too low? Just adding the "failcnt" value? Adding 10? Adding 100? Multiplying with 2? Multiplying with 5?

For example, I have privvmpages 57536:71632. Then, after some time failcnt=100. Adjusting to $((57536+100)):$((71632+100)) seems not to help, so I did $((57536+10000)):$((71632+10000)) = 67536:61632.

I hope you can follow me now. The key question is: How to change the limit values for a given "failcnt" in order to ensure reliable VEs?

Quote:
This might be a valid behavior. Here is a nice article.
http://wiki.openvz.org/UBC_failcnt_reset

Thank you. I will wait the 5 minutes after rebooting the VE As said, I just want to do this in my "learning phase".

divB

Subject: Re: Disabling all limits
Posted by maratrus on Fri, 10 Apr 2009 10:31:08 GMT
View Forum Message <> Reply to Message

Quote:
I think we have a misunderstanding here. In my case, the goal is to find out how much ressources are needed by my servers

The possible solution is to disable all limits at all. Wait for a "learning phase" and after that learn the "maxheld" column of /proc/user_beancounters output.

Quote:
So in this case I want each VE to give as much ressources as they need.

Why do you want to obtain the precise limits boundary then?
As far as I understand your scenario, you always increase barrier:limit values in case of failure.
Why don't you switch off all limits?

Quote:
Just adding the "failcnt" value?

failcnt = failcounter. It's just a counter. So, if your application bumped into memory shortage and was refused to allocate virtual memory, failcounter increased by 1.

Subject: Re: Disabling all limits
Posted by divB on Fri, 10 Apr 2009 15:33:57 GMT
View Forum Message <> Reply to Message

Hi again,

And thank you for this hint, I think I will do that.

The "desaster" happened the first time today: The VE with the vital LDAP server had (obviously) too few ressources causing the slapd to be killed. And all other services including logins were unavailable.

But this brings me back to the original question: How to turn them off completely? I guess I could just take the values of VE0; is this a good idea?