Subject: extreme load during high disk i/o
Posted by eatingorange on Fri, 31 Aug 2007 16:02:37 GMT
View Forum Message <> Reply to Message

Greetings all,

I'm curious if there is a known issue with respect to load averages during periods of high disk i/o. I have a hostnode (dual Xeon 3.0, 8gb RAM, 2x72gb U320 16mb cache (soft raid1)) that's been running fine for about a month. Yesterday I tried to import some tables from a MySQL dump file and brought the entire server to it's knees. I had to set cpulimit to 1% on the ve and still seeing load average of about 4.5 on the hostnode during this import. I've tweaked around with memory allocation to ve and tried adjusting ioprio, and ensured that write cache is enabled on the drives but the only way I can even slightly make this work is by severely limiting cpu availability to the ve and still seeing fairly high loadavg on the machine.

At one point, I had a loadavg of 89 on this very competent server.

Kernel: 2.6.18-8.1.8.el5.028stab039.1PAE on CentOS 5.

Thank you for your thoughts on this.

Eric

Subject: Re: extreme load during high disk i/o
Posted by kir on Fri, 31 Aug 2007 16:13:26 GMT
View Forum Message <> Reply to Message

High loadavg during high I/O activity periods just shows that there are that many processes waiting for the disk I/O. Sometimes it's normal, sometimes it's not. Definitely it's not normal then other VEs starts to suffer because of high I/O activity in one VE.

Well, lowering ioprio on this offensive VE should help; setting cpulimit should also ease the problem, so you did it right.

Other than that there's no cure, not counting RAM upgrade (to have more disk cache), faster and/or RAIDed disks.

Subject: Re: extreme load during high disk i/o
Posted by dowdle on Fri, 31 Aug 2007 21:10:29 GMT
View Forum Message <> Reply to Message

I wouldn't call a 4.5 load an "extreme load". I would also guess that the host machine and VPSes were still quite responsive.

My point is, don't use that single indicator as proof of a problem.

I realize you had a considerable amount of memory (8GB) so I don't know if a RAM upgrade would help any... assuming the machine in question can take more RAM than that.

Just out of curiousity, how big was your MySQL dump file and approximately how long did the operation take?

Subject: Re: extreme load during high disk i/o
Posted by eatingorange on Fri, 31 Aug 2007 21:34:00 GMT
View Forum Message <> Reply to Message

Hi,

Actually 4.5 is where the machine has "settled in" now that CPU on the VE in question is limited to 1%. When I first started the import, load average on the hostnode was 89 and climbing (that's what I was referring to when I said "extreme load"). It did in fact cause problems on the other VEs because their disk i/o got bound up as well. Loads on the other machines got high enough for Sendmail to refuse connections, which is when I realized I had a real problem on my hands.

The MySQL file is 10gb and contains about 28 million records. It's been running for about 18 hours now and it's 20 percent complete. I've tried the --unbuffered option during mysql import and it makes no difference. The thing just spirals completely out of control unless cpu availability is severely limited on the VE.

Thanks

Eric

Subject: Re: extreme load during high disk i/o
Posted by dowdle on Fri, 31 Aug 2007 21:46:22 GMT
View Forum Message <> Reply to Message

Wow... not that *IS* a high load... but given the fact that your importing 10GB of MySQL data... I wouldn't really expect anything different.

In the future, if possible, you could migrate the VPS in need of the MySQL import... to another machine without any VPSes on it, do your import without hurting anything... and when it is done, migrate it back. Yeah, that's a lot of work and I'm sure it would take a bit of time to do the transfer to the destination host node ... but less time back if you "-r no" in the first place.

Can you renice the process?

Subject: Re: extreme load during high disk i/o
Posted by dowdle on Fri, 31 Aug 2007 21:48:03 GMT
View Forum Message <> Reply to Message

Oh, since this is a disk I/O issue as well as a CPU issue... isn't there a way to lower the disk I/O priority of the VPS?

Subject: Re: extreme load during high disk i/o
Posted by eatingorange on Fri, 31 Aug 2007 22:08:19 GMT
View Forum Message <> Reply to Message

I tried renice but since it's not a cpu contention issue, it really had no effect. And I did set --ioprio to 1 and then to 0 and that didn't change things much either. What really freaked me out is that this had such a visible effect on the other VE, which was running Sendmail and having a very hard time writing its maillog to the point that load started creeping to scary levels on that VE as well. Maybe I've just never noticed how much iowait 'mysql < dump.sql' can generate. Left unchecked, I fear this thing would have just ground to a halt.

Subject: Re: extreme load during high disk i/o
Posted by locutius on Sat, 01 Sep 2007 13:13:21 GMT
View Forum Message <> Reply to Message

doesnt help you are running soft raid. your poor box must contend with calculating parity while it swaps RAM because of the size of the dump and maintain a level of service for any other disk bound services e.g. sendmail

breaking that dump into 2GB pieces will relieve the pressure on disks and i predict you will see significant reduction in job time due to reduction in errors and inefficiences caused by the current thrash

even better, move to hardware raid 1

Subject: Re: extreme load during high disk i/o
Posted by dowdle on Sun, 02 Sep 2007 16:32:45 GMT
View Forum Message <> Reply to Message

There are few issues here:

1) VE / Host Isolation - OpenVZ isolates VEs from each other and the host node in such a way as not to allow (a properly configured) VE to hurt other VEs or the host node. In this case OpenVZ's isolation isn't working as well as we would hope because the VE is having a big impact on other VEs and the host node.

2) VE Performance - Should you do things inside of a VE that is ill-advised even on a host node? To clarify, chances are you have configured your VE to have less resources than your host node making the situation worse than doing your operation on a host node... so to rephrase... Should you do things inside of a VE that is ill-advised even on a host node with more resources than the VE?

3) I don't want tell you that you shouldn't use software (Linux kernel) based RAID on your host node as I've read that in some cases it out performs some hardware RAID... but then again... in this case, the software RAID part might indeed be a contributing factor. At this point I don't know for sure.

I have a few questions.

If you were to do an import of a 10GB mysqldump file on a physical host configured with the same amount of RAM you have allocated to your VE... what impact would that have on the physical machine? How much I/O wait would you encounter and how large would the load go... and how long would the job take? With that information, it would be nice to compare that to the performance of your VE to see if there is indeed a VE performance issue. Of course, the steps you took to address #1 above, only degrade the performance further.

I'm guessing, without much information to back it up, that the operation in question wouldn't perform very well anywhere... and that as a responsible system admin, you should strongly consider breaking up the mysqldump import into smaller chunks... as locutius suggested... as that strategy alone may resolve the symptoms... at least regarding the VE Performance part.

I'm kinda in the dark here as I have not tried to import a 10GB mysqldump file and have no idea if it is a piece of cake or a major pain.

Regarding your suggestion that because you are using OpenVZ that you don't have to worry about your host node nor about anything your users do in a VE... and that you are lucky because you were paying attention THIS time... I disagree. If you are in charge of the host node, you are also (to some degree) responsible for what goes on in the VEs. You should always try to monitor your systems and be proactive to avoid problems. This isn't something OpenVZ is going to solve for you no matter how well it works.

If you (or your VE root users) don't keep your VE updated with security patches... or users set bad passwords... and your VE gets hacked and becomes a spambot node... who is going to feel the pain? The same goes for performance issues.

We'll continue to try to figure out what the problems are here and how to solve them... and do appreciate your willingness to provide information when asked. While this operation is probably not something that should be considered a common problem, resolving it will definitely improve