Subject: Heavy Disk IO from a single VM can block the other VMs on the same host
Posted by Hubert Krause on Tue, 29 Nov 2011 16:13:26 GMT
View Forum Message <> Reply to Message

Hello,

my environment is a Debian squeeze host with a few debian squeeze guests. The private and root filesystems of the guest are locatet on the same raid device (raid5) in an luksCrypt Container in an LVM container on an ext4 partition with nodelalloc as mountoption. If I run the tool stress:

stress --io 5 --hdd 5 --timeout 60s (which means fork 5 threads doing read/write access and 5 threads doing constantly fsync) the responsivness of the other VMs is very bad. That means, Isolation for IO operations is not given. I've tried to reduce the impact of the VM with 'vzctl set VID --ioprio=0'. There was only a minor effect, my application on the other VM where still not responsive.

Any Idea how to prevent a single VM to disturb the other VMs regarding diskIO?

Greetings

Hubert

Subject: Re: Heavy Disk IO from a single VM can block the other VMs on the same host
Posted by Bogdan-Stefan Rotariu on Thu, 01 Dec 2011 15:49:19 GMT
View Forum Message <> Reply to Message

On Nov 29, 2011, at 18:13, Hubert Krause <hubert.krause@inform-software.com> (by way of HubertKrause <hubert.krause@inform-software.com>) (by way of HubertKrause <hubert.krause@inform-software.com>) wrote:

> Hello,
>
> my environment is a Debian squeeze host with a few debian squeeze
> guests. The private and root filesystems of the guest are locatet on
> the same raid device (raid5)

maybe offtopic, maybe not, but stop using raid5 for VM deployment, use raid10, raid1, raid0 -- with lvm and snapshots

raid5 will always be slow on io, as it has checksums because "recalculation and redistribution of parity data on a per-write basis"

Subject: Re: Heavy Disk IO from a single VM can block the other VMs on the same host
Posted by Kirill Korotaev on Thu, 01 Dec 2011 17:27:49 GMT
View Forum Message <> Reply to Message

That's most likely due to a single file system used for containers - journal becomes a bottleneck. fsync forces journal flushes and other workloads begin to wait for journal... In reality workload looks like this are typical for
heavy loaded databases or mail systems only.

How to improve:
- increase journal size
- split file systems, i.e. run each container from it's own file system

Thanks,
Kirill

On Nov 29, 2011, at 20:13 , Hubert Krause wrote:

> Hello,
>
> my environment is a Debian squeeze host with a few debian squeeze
> guests. The private and root filesystems of the guest are locatet on
> the same raid device (raid5) in an luksCrypt Container in an LVM
> container on an ext4 partition with nodelalloc as mountoption. If I run
> the tool stress:
>
> stress --io 5 --hdd 5 --timeout 60s (which means fork 5 threads doing
> read/write access and 5 threads doing constantly fsync) the
> responsivness of the other VMs is very bad. That means, Isolation for
> IO operations is not given. I've tried to reduce the impact of the
> VM with 'vzctl set VID --ioprio=0'. There was only a
> minor effect, my application on the other VM where still not
> responsive.
>
> Any Idea how to prevent a single VM to disturb the other VMs regarding
> diskIO?
>
> Greetings
>
> Hubert

Subject: Re: Heavy Disk IO from a single VM can block the other VMs on the same host
Posted by quantact-tim on Fri, 02 Dec 2011 18:18:18 GMT
View Forum Message <> Reply to Message

You can use vzctl --ioprio to set relative disk I/O priorities:
http://wiki.openvz.org/I/O_priorities_for_VE

-Tim

--
Timothy Doyle
CEO
Quantact Hosting Solutions, Inc.
tim@quantact.com
http://www.quantact.com

On 12/01/2011 09:27 AM, Kirill Korotaev wrote:
> That's most likely due to a single file system used for containers - journal becomes a bottleneck.
> fsync forces journal flushes and other workloads begin to wait for journal... In reality workload looks like this are typical for
> heavy loaded databases or mail systems only.
>
> How to improve:
> - increase journal size
> - split file systems, i.e. run each container from it's own file system
>
> Thanks,
> Kirill
>
>
> On Nov 29, 2011, at 20:13 , Hubert Krause wrote:
>
>> Hello,
>>
>> my environment is a Debian squeeze host with a few debian squeeze
>> guests. The private and root filesystems of the guest are locatet on
>> the same raid device (raid5) in an luksCrypt Container in an LVM
>> container on an ext4 partition with nodelalloc as mountoption. If I run
>> the tool stress:
>>
>> stress --io 5 --hdd 5 --timeout 60s (which means fork 5 threads doing
>> read/write access and 5 threads doing constantly fsync) the
>> responsivness of the other VMs is very bad. That means, Isolation for
>> IO operations is not given. I've tried to reduce the impact of the
>> VM with 'vzctl set VID --ioprio=0'. There was only a
>> minor effect, my application on the other VM where still not
>> responsive.
>>
>> Any Idea how to prevent a single VM to disturb the other VMs regarding
>> diskIO?
>>

>> Greetings
>>
>> Hubert

Subject: Re: Heavy Disk IO from a single VM can block the other VMs on the same host
Posted by Hubert Krause on Tue, 06 Dec 2011 17:18:08 GMT
View Forum Message <> Reply to Message

Hello Kirill,

Am Thu, 1 Dec 2011 21:27:49 +0400
schrieb Kirill Korotaev <dev@parallels.com>:

> That's most likely due to a single file system used for containers -
> journal becomes a bottleneck. fsync forces journal flushes and other
> workloads begin to wait for journal... In reality workload looks like
> this are typical for heavy loaded databases or mail systems only.
>
> How to improve:
> - increase journal size
> - split file systems, i.e. run each container from it's own file
> system

I've created another lv with an ext4 filesystem with maximum
journal-size and mounted this filesystem
under /var/lib/vz/private/<VID>. I will call this vm as VM-sep. All
other vhosts where kept inside the volume as before. Than I
start stressing the VM-sep and tested the impact to the other VMs. It
was exactly the same as if I run all VMs on the same partition.

There was indeed a difference, when I stress the Host itself. If I do
filesystem stress in the same Partition (/var/lib/vz) The perfomance of
VM is much worse (similar to stress in a VM, a little better) than if I
would stress in a completly different Partition (/var/tmp in my case)

To get some Numbers: (not very sientific, but good for a measure)

Throughput of a Webserver in a VM called VM-web in KB/s:
* without stress 101.9
* stress /var/tmp on host 24.3
* stress /var/lib/vz on Host 10.5
* stress a vm, not VM-web, same fs 8.3
* stress VM-sep 7.6

Maybe the Diskencryption plays a role, maybe there is something in the
VM-Isolation layer, I have no clue.

But as you mentione before this workload is typical for heavy
loaded databases or mail systems only. Neither of this application will
run in my vm-environment. So I will ignore this.

Greetings,

Hubert

> On Nov 29, 2011, at 20:13 , Hubert Krause wrote:
> > my environment is a Debian squeeze host with a few debian squeeze
> > guests. The private and root filesystems of the guest are locatet on
> > the same raid device (raid5) in an luksCrypt Container in an LVM
> > container on an ext4 partition with nodelalloc as mountoption. If I
> > run the tool stress:
> >
> > stress --io 5 --hdd 5 --timeout 60s (which means fork 5 threads
> > doing read/write access and 5 threads doing constantly fsync) the
> > responsivness of the other VMs is very bad. That means, Isolation
> > for IO operations is not given. I've tried to reduce the impact of
> > the VM with 'vzctl set VID --ioprio=0'. There was only a
> > minor effect, my application on the other VM where still not
> > responsive.
> >
> > Any Idea how to prevent a single VM to disturb the other VMs
> > regarding diskIO?

--
Dr. Hubert Krause
Geschäftsbereich Risk & Fraud
INFORM GmbH, Pascalstr. 23, 52076 Aachen, Germany
Tel. (+49)2408 9456 5145
e-mail: hubert.krause@inform-software.com,
http://www.inform-software.de/ INFORM Institut fur Operations Research
und Management GmbH Registered AmtsG Aachen HRB1144 Gfhr.Adrian Weiler