DDN STORAGE

TECHNICAL SUPPORT BULLETIN

DDN TECHNICAL SUPPORT | SPT-TSB-0095 (A1) | 22 MAY 2019

EXAScaler™

Virtio Issue Slows File System Performance

ALERT! On embedded EXAScaler systems running in a declustered RAID environment (SFA OS 11.1 and higher), a memory fragmentation issue affecting the `virtio_scsi` driver can degrade EXAScaler performance and cause the file system to become unresponsive.

Issue Summary

Virtio is an industry-standard virtualization technology for high-speed, scalable I/O transport among virtual machines and virtualized storage, network, and I/O devices. SFA OS adopted this technology as part of the design transition to declustered RAID in SFA OS 11.1 and higher. EXAScaler added support for virtio with its support for SFA OS 11.1 in EXAScaler 4.0.

EXAScaler solutions running on embedded VMs under SFA OS 11.1 and higher can encounter issues with virtio when serving a large number of native Lustre clients. When many Lustre clients are accessing Lustre OST targets, the OSS server will rapidly allocate and deallocate network buffer memory to handle client I/O. This can cause server memory to become fragmented. When this occurs, the `virtio_scsi` driver has difficulty allocating contiguous memory regions. If allocation fails, `virtio_scsi` reports a "page allocation failure" error and dumps stack traces to the Linux syslog. If the issue persists, it can impact system performance and cause the file system to become unresponsive.

Affected Systems

Only systems with ***all*** the following conditions are affected by this issue:

- EXAScaler 4.0 and higher
- SFA OS 11.1 and higher
- EXAScaler executes on an embedded VM on an SFA14KXE (ES14K) or SFA7990E (ES7990)

EXAScaler versions earlier than 4.0 are ***not*** affected. EXAScaler 4.x versions running under legacy SFA OS versions (3.x) or on the SFA7700XE (ES7K) are also ***not*** affected. EXAScaler 4.x versions running on external servers are similarly ***not*** affected.

The `virtio_scsi` driver is part of the `kmod-virtio-scsi` RPM package.

Workaround

The recommended workround for this issue is setting the `sysctl` parameter:

```
vm.min_free_kbytes: 1048576
```

This setting will "raise the floor" and cause the Linux kernel to retain more memory regions in order to accommodate more `virtio_scsi` driver requests for contiguous buffer allocations.

To implement the workaround:

Step 1. Back up the current EXAScaler configuration file `/etc/ddn/exascaler.conf`.

Step 2. Back up the current `sysctl` configuration file `/etc/sysctl.conf`.

Step 3. With a text editor, edit the EXAScaler configuration file `/etc/ddn/exascaler.conf` as follows:

a. Find the **`[sysctl_defaults]`** section heading.

b. Under the **`[sysctl_defaults]`** heading, add the entry:

```
vm.min_free_kbytes: 1048576
```

Step 4. Copy the updated configuration file to all EXAScaler servers with the command:

```
$ sync-file /etc/ddn/exascaler.conf
```

Step 5. Install the copied **`sysctl`** settings on all EXAScaler servers with the command:

```
$ clush -a "es_install --yes --steps os"
```

Step 6. Enable the new settings using the **`sysctl`** command:

```
$ clush -a “sysctl -p /etc/sysctl.conf”
```

Please note that the workaround reduces the number of page allocation errors, but it will not eliminate them completely.

Resolution

A resolution for the issue is expected in a future Linux distribution.

Contacting DDN Technical Support

Please contact DDN Technical Support at any time if you have questions or require assistance. Support can be reached by phone, by email, or on the web as listed below.

Web	
DDN Community Support Portal	https://community.ddn.com/login
Portal Assistance	webportal.support@ddn.com
Telephone	
DDN Support Worldwide Directory	https://www.ddn.com/support/global-services-overview/
Email	
Support Email	support@ddn.com
Bulletins	
Support Bulletins	http://www.ddn.com/support/technical-support-bulletins
End-of-Life Notices	http://www.ddn.com/support/end-of-life-notices
Bulletin Subscription Requests	support-tsb@ddn.com