AIX FC Performance improvements for IBM AIX FC and FCoE device driver stacks

Article Number: 660 | Rating: Unrated | Last Updated: Fri, Jan 31, 2020 6:00 PM

Performance improvements for IBM AIX FC and FCoE device driver stacks

Multiple I/O queue support

Yadagiri Rajaboina, Kiran Anumalasetty, Vinod Kumar Boddukuri, and Prashantha Subbarao
Published on January 06, 2017

13

This article describes performance improvements for IBM® AIX® Fibre Channel (FC) / Fibre Channel over Ethernet (FCoE) device drivers stack for 16 Gb FC (Feature Code: EN0A) and10 Gb FCoE (Feature Code: EN0H) HBAs. The AIX FC driver stack includes an initiator mode Small Computer System Interface (SCSI) protocol driver and an adapter driver. The existing AIX FC adapter driver maintains a SCSI I/O queue for submitting all I/O requests to the FC HBA. Scaling issues have been observed with the existing FC stack with extreme I/O transactions per second (IOPS) and small I/O sizes. This is due to I/O serialization through a single I/O queue.

To improve the number of IOPS on smaller block size I/O requests, the multiqueue functionality is introduced with the 16 Gb FC or 10 Gb FCoE adapter driver starting from:

- AIX releases – AIX 7.2 TL01 SP1, AIX 7.1 TL04 SP3 and AIX 6.1 TL09 SP8
- VIOS release – VIOS 2.2.4.30 and VIOS 2.2.5.0

Figure 1 depicts how I/O is parallelized on multiple I/O queues with the improvements discussed in this article.

Figure 1. A traditional FC driver stack versus an improved FC driver stack

Configuration details

The following configuration is used for the performance analysis of random read operations with a block size of 4 KB

- IBM Power® System E870 server with 64 processors at a frequency of 4.350 GHz
- IBM FlashSystem® 900 with eight storage FC ports:
- With FC
    - Brocade 16 Gb FC Switch: 2498-B24
    - PCIe2 two-port 16 Gb FC adapter (Feature Code: EN0A)
- With FCoE:
    - PCIe2 10 Gb four-port FCoE adapter (Feature Code: EN0H)
    - Brocade 10 Gb FCoE Switch
- With Native (stand-alone) AIX configuration:
    - Operating system: AIX 7.2 TL01 SP1
    - Number of processors: 32
- With N_ Port ID Virtualization (NPIV) configuration:
    - VIOS release: 2.2.5.0
    - NPIV client OS: AIX 7.2 TL01 SP1
    - Number of processors on VIOS host: 32
    - Number of processors on each NPIV client: 4

Implementation details

To support multiple I/O queue feature, a new Object Data Manager (ODM) attribute, num_io_queues, is introduced for FC/FCoE devices (fcs) to indicate the number of I/O queues configured in the FC adapter driver. Each I/O queue is associated with a hardware work queue in the FC HBA. All the I/O requests issued to a particular hdisk will be mapped to the same SCSI I/O queue. Each SCSI I/O queue can service multiple hdisks, However, I/O request to a given hdisk cannot be distributed to multiple SCSI I/O queues.

Example: For a 16 Gb FC HBA, the ODM stanza for num_io_queues attribute is shown below:

```
# lsdev | grep fcs
fcs0       Available 00-00       PCIe2 2-Port 16Gb FC Adapter (df1000e21410f103)
fcs1       Available 00-01       PCIe2 2-Port 16Gb FC Adapter (df1000e21410f103)

# odmget -q name=fcs0 CuDv
CuDv:
        name = "fcs0"
        status = 1
        chgstatus = 0
        ddins = "pci/emfcdd"
        location = "00-00"
        parent = "pci0"
        connwhere = "0"
        PdDvLn = "adapter/pciex/df1000e21410f10"
#
# odmget -q uniquetype="adapter/pciex/df1000e21410f10" PdAt | grep -p num_io_queues
PdAt:
        uniquetype = "adapter/pciex/df1000e21410f10"
        attribute = "num_io_queues"
```

```
        deflt = "8"

        values = "1-16,1"

        width = ""

        type = "R"

        generic = "DU"

        rep = "nr"

        nls_index = 67

#
```

The value of this attribute can be changed using the chdev command or the System Management Interface Tool (SMIT) interface. The possible values are:

```
# lsattr -l fcs0 -a num_io_queues -R

1...16 (+1)
```

To enable multiple I/O queues, the HBA's direct memory access (DMA) resources should be sufficient to distribute I/O requests across multiple queues. The existing ODM attribute, io_dma, controls the amount of I/O DMA region that the adapter driver requests while configuring the HBA.

Default ODM attribute values

This section provides the default values for the ODM attributes related to the multiple I/O queue feature.

For AIX 7.2 TL01 SP1 and VIOS 2.2.5 releases

The default value of the num_io_queues attribute is set to 8 and to have sufficient DMA resources, the default value of the io_dma attribute is increased to 256, starting with AIX 7.2 TL01 SP1 and VIOS 2.2.5 releases.

```
# lsattr -El fcs0 | grep -e num_io_queues -e io_dma

io_dma          256          IO_DMA                            True

num_io_queues   8            Desired number of IO queues       True
```

For AIX 7.1 TL04 SP3, AIX 6.1 TL09 SP8 and VIOS 2.2.4.30 releases

The default value of the num_io_queues attribute is set to 1 and io_dma is set to 64 for the AIX 7.1 TL04 SP3, AIX 6.1 TL09 SP8 and VIOS 2.2.4.30 releases.

```
# lsattr -El fcs0 | grep -e num_io_queues -e io_dma

io_dma          64         IO_DMA                            True
```

```
num_io_queues   1         Desired number of IO queues              True
```

As mentioned earlier in this article, there should be sufficient DMA resources to enable support for multiple I/O queues. Therefore, the value of the io_dma attribute should be increased from 64 to 256. In case the user changes only the num_io_queues' value without increasing the io_dma value to 256, the adapter instance will be configured with a single SCSI I/O queue and the following informational error will be logged with the AIX error log.

```
# errpt | grep fcs0

29FA8C20   0629173616 I O fcs0  Additional FC Adapter Information
```

Steps for tuning the num_io_queues attributes using the chdev command

You need to perform the following steps to tune the num_io_queues attributes using the chdev command for 16 Gb FC HBAs:

- Unconfigure the device instance.

```
#rmdev -Rl fcs0
```

- Change the attribute to the required value (say 16).

```
# chdev -l fcs0 -a num_io_queues=16

fcs0 changed
```

- Configure the device instance.

```
# cfgmgr -l fcs0
```

- Verify if the attribute is set to the required value.

```
#  lsdev | grep fcs0

fcs0       Available 00-00       PCIe2 2-Port 16Gb FC Adapter (df1000e21410f103)

# lsattr -El fcs0 | grep num_io_queues

num_io_queues   16         Desired number of IO queues                    True
```

Performance results – random read operations with a block size of 4 KB

The following results are for AIX native [that is, physical HBAs owned by the logical partition (LPAR)] case. I/O requests are running in parallel on the FlashSystem 900 storage targets using the default shortest_queue algorithm on the hdisk devices.

Figure 2. AIX (native) results for 16 Gb FC HBA

Figure 3. AIX (native) results for 10 Gb FCoE HBA

The following results are for the NPIV (that is, VIOS owning physical HBAs and num_io_queues tuned on VIOS) case. I/O requests are running in parallel on all the I/O paths for a given disk on each NPIV client using the default shortest_queue algorithm on the Flash System 900 storage disks.

Figure 4. NPIV results for 16 Gb FC and 10 Gb FCoE HBA

Conclusion

In native configuration, the number of IOPS for random read operations with a block size of 4KB, for a single FC HBA port case increased by approximately 2.5 times with the improved FC stack, which is a significant improvement. The achieved IOPS count of 390,000 is very close to the line speed for a single FC HBA port.

The IOPS gain for random read operations with a block size of 4 KB in the NPIV configuration is almost equivalent to that of the native configuration when the number of clients is six or more.