Hewlett Packard Enterprise

How to calculate suitable LUN queue depths

Host I/O queues and HPE 3PAR StoreServ Storage

Contents

Technical white paper

Introduction

This white paper discusses the approach for tuning the host I/O queue depth limits in a typical data center environment consisting of multiple servers connected to a Hewlett Packard Enterprise 3PAR StoreServ system via a storage area network (SAN).

The intended audience for this white paper includes server administrators and storage administrators.

I/O queuing

To improve I/O performance, hosts can simultaneously submit more than one I/O to a disk at a time. This can be due to multiple processes concurrently accessing a file system or applications using asynchronous I/O. Generating simultaneous I/Os is especially important with RAID array subsystems where a virtual disk is backed by multiple physical disks. In such a situation, submitting multiple I/Os allows the array to access multiple physical disks and, therefore, achieves more IOPS than would have been possible by submitting one I/O at a time.

If an HPE 3PAR StoreServ Storage system port is connected to many hosts through a fabric, it is possible that if the hosts simultaneously do a large amount of I/O, the target port could run out of I/O buffers. This results in the target port rejecting any new incoming I/O requests and issuing a TASK_SET_FULL SCSI status message (also known as QUEUE FULL) until at least one outstanding request has completed, freeing up I/O buffers. Hosts receiving the QUEUE FULL status can retry their rejected commands but performance degradation can still occur as the number of host retry operations increases.

To prevent this condition, you can tune the I/O queue depth on the hosts so they are unable to saturate the target ports.

Calculating queue depths

With the wide-striping architecture of a HPE 3PAR StoreServ system, it is important to balance the load as evenly as possible across all available nodes and ports. A system configuration that has nodes with substantially different loads may not deliver the full potential performance of the hardware.

HPE recommends that the total aggregate of the maximum number of outstanding SCSI commands from all hosts connected to a port of a HPE 3PAR StoreServ Storage system should be less than the maximum queue depth of the port.

This recommendation can be represented by the following equation:

$$\text{Port-QD} \geqslant \text{Host1}(P * L * QD) + \text{Host2}(P * L * QD) + \ldots + \text{Hostn}(P * L * QD)$$

Where,

Port-QD = Maximum queue depth of the array target port

P = Number of host paths connected to the array target port

L = Number of LUNs presented to the host via the array target port

QD = LUN queue depth on the host

The maximum queue depth of the array target port (Port-QD), the number of hosts and the number of host paths (P) are essentially fixed values so this leaves the number of LUNs presented (L) and host LUN queue depth (QD) as variables.

The host LUN queue depth is the maximum number of I/O operations that can be run in parallel to a device. It is an operating system parameter that is set either in the HBA driver or the SCSI subsystem. The default value varies depending on the operating system but it is usually in the range of 16–32.

If the number of hosts and LUNs are low, then the total aggregate of the maximum number of outstanding SCSI commands are much lower than the maximum queue depth of the array target port from all hosts so the host LUN queue depths can be increased to improve the performance potential. However, if there are a large number of hosts and LUNs, then the maximum LUN queue depths on the hosts should be reduced but performance degradation also occurs if the values are set too low.

Although the most widely used method of controlling the array port queue depth is to change the host LUN queue depth, storage administrators should carefully consider the number of hosts connected to an HPE 3PAR StoreServ port and the number of LUN exports for calculating the throttling configuration values.

This is particularly important if the storage layout has a lot of small Virtual Logical Unit Numbers (VLUNs) as a result of a migration from another type of array as it may not be optimal for the wide striping architecture of HPE 3PAR StoreServ. This means presenting fewer and larger LUNs to the operating system.

HPE 3PAR StoreServ Storage queue limits

Queue limits per HBA

On HPE 3PAR StoreServ Storage, the Host Bus Adapters (HBAs) have multiple ports and the I/O queue depth limits are per port.

The I/O queue depth of each HPE 3PAR StoreServ HBA model is as follows:

HBA	PROTOCOL	ARRAY	BUS	SPEED	PORTS	MAX. QUEUE DEPTH
Emulex LP11002	FC	F200, F400, T400, T800	PCI-X	4 Gbps	2	959
HPE 3PAR FC044X	FC	F200, F400, T400, T800	PCI-X	4 Gbps	4	1638
QLogic QLA4052C	iSCSI	F200, F400, T400, T800	PCI-X	1 Gbps	2	512
Emulex LPe12002	FC	StoreServ 7000	PCIe	8 Gbps	2	3276
Emulex LPe12004	FC	StoreServ 7000, StoreServ 10000	PCIe	8 Gbps	4	3276
QLogic QLE8242	FCoE	StoreServ 7000, StoreServ 10000	PCIe	10 Gbps	2	1748
QLogic QLE8242	iSCSI	StoreServ 7000, StoreServ 10000	PCIe	10 Gbps	2	2048
QLogic QTH8362	FCoE	StoreServ 8000, Storeserv 20000	PCIe	10 Gbps	2	1748
QLogic QTH8362	iSCSI	StoreServ 8000, Storeserv 20000	PCIe	10 Gbps	2	2048
Emulex LPe16002	FC	StoreServ 7000, StoreServ 8000, StoreServ 10000	PCIe	16 Gbps	2	3072
Emulex LPe16004	FC	StoreServ 8000, StoreServ 20000	PCIe	16 Gbps	4	3072

Determining the HBA type

There are two different CLI commands that can be used to determine which HBA models the HPE 3PAR StoreServ Storage is using. The `shownode` command lists the PCI adapters installed in each node:

```
cli% shownode -pci
Node Slot Type  -Service_LED-  -Manufacturer- -Model--  --Serial--  -Rev-  Firmware
   0    0   FC  Off            EMULEX         LPe12004  BT01477194  03     2.00.X.14
   0    1   FC  Off            EMULEX         LPe12004  BT95188340  03     2.00.X.14
```

The showport command lists the HBA associated with each port:

```
cli% showport -i
N:S:P         Brand         Model         Rev    Firmware      Serial        HWType
0:0:1         EMULEX        LPe12004      03     2.00.X.14     BT01477194    FC
0:0:2         EMULEX        LPe12004      03     2.00.X.14     BT01477194    FC
0:0:3         EMULEX        LPe12004      03     2.00.X.14     BT01477194    FC
0:0:4         EMULEX        LPe12004      03     2.00.X.14     BT01477194    FC
0:1:1         EMULEX        LPe12004      03     2.00.X.14     BT95188340    FC
0:1:2         EMULEX        LPe12004      03     2.00.X.14     BT95188340    FC
0:1:3         EMULEX        LPe12004      03     2.00.X.14     BT95188340    FC
0:1:4         EMULEX        LPe12004      03     2.00.X.14     BT95188340    FC
...
```

Use the HBA model name from the output of either command in conjunction with the previous table to determine the Port-QD value for the HPE 3PAR StoreServ Storage ports.

Ports and VLUNs

Virtual volumes are made available to one or more hosts by creating a virtual volume (VV)-LUN pairing between a virtual volume and a LUN. The createvlun command creates a VLUN template that enables export of a VV as a VLUN to one or more hosts. An active VLUN is created when the current system state matches the rule established by the VLUN template.

Use the showvlun command to display the active VLUNs. Using the -nodes, -slots and -ports switches, you can show all the hosts and the volumes exported to them for the ports of a HBA. This is ideal for working out the maximum queue depth loads on a HBA port:

```
cli% showvlun -a -nodes 0 -slots 1 -ports 1
Domain  Lun VVName           HostName -Host_WWN/iSCSI_Name-  Port     Type
-         0 h105c011_vv0     h105c011 50014280029C4E24      0:1:1     host
-         1 h105c011_vv1     h105c011 50014280029C4E24      0:1:1     host
-         2 h105c011_vv2     h105c011 50014280029C4E24      0:1:1     host
-         3 h105c011_vv3     h105c011 50014280029C4E24      0:1:1     host
-         4 h105c002_vv0     h105c002 50014380029C4EF4      0:1:1     host
-         5 h105c002_vv1     h105c002 50014380029C4EF4      0:1:1     host
-         6 h105c002_vv2     h105c002 50014380029C4EF4      0:1:1     host
-         7 h105c002_vv3     h105c002 50014380029C4EF4      0:1:1     host
-         8 h105c010_vv0     h105c010 50014280029C4DA8      0:1:1     host
-         9 h105c010_vv1     h105c010 50014280029C4DA8      0:1:1     host
```

```
-          0 h105c010_vv2     h105c010 50014280029C4DA8      0:1:1     host
-          1 h105c010_vv3     h105c010 50014280029C4DA8      0:1:1     host
-          2 h105c014_vv0     h105c014 50014380029C4D70      0:1:1     host
-          3 h105c014_vv1     h105c014 50014380029C4D70      0:1:1     host
-          4 h105c014_vv2     h105c014 50014380029C4D70      0:1:1     host
-          5 h105c014_vv3     h105c014 50014380029C4D70      0:1:1     host
-          6 h105c005_vv0     h105c005 50014380029C4E54      0:1:1     host
-          7 h105c005_vv1     h105c005 50014380029C4E54      0:1:1     host
-          8 h105c005_vv2     h105c005 50014380029C4E54      0:1:1     host
-          9 h105c005_vv3     h105c005 50014380029C4E54      0:1:1     host
------------------------------------------------------------------------------
          20 total
```

The output shows that port 1 of this HBA has five hosts connected to it and each host has four VLUNs presented making a total of 20 VLUNs on this port.

Queue depth calculation example

Using the output from the CLI commands above as the basis of our example, the HPE 3PAR StoreServ Storage in question has an Emulex LPe12004 4-port HBA in node 0 slot 1, which has a maximum queue depth of 3276 per port and there are a total of 20 VLUNs presented through port 1.

If the hosts all have a LUN queue depth of 32, then the queue depth equation:

$$Port\text{-}QD \geqslant Host1(P * L * QD) + Host2(P * L * QD) + \ldots + Hostn(P * L * QD)$$

becomes

$$3276 \geqslant 640$$

Therefore, the maximum number of host I/Os is well below the HBA queue limit. In this scenario, one could consider raising the host LUN queue depth to 64 or more to potentially increase the maximum throughput.

However, a host would only benefit from having a larger LUN queue depth if the applications running can issue sufficient simultaneous I/Os. Use the resource monitor on Microsoft® Windows® operating systems or iostat on UNIX® or Linux® to monitor the device queue depths. Only if the device queue depths are able to reach the per LUN queue depth limit would it be worthwhile raising the limit.

Monitoring port queues

In the previous example, if there were 20 hosts each with 10 VLUNs presented then the maximum number of host I/Os possible would be 6400, which is almost double the HBA queue limit. Despite the recommendation, it is likely that not all hosts use all their queue depth at the same time. In this situation, you should monitoring the queues of the HPE 3PAR StoreServ Storage ports to check if the HBA queue limit is being exceeded.

You can monitor the port queue depths with the statport command. Use the -nodes and -slots switches to show all the ports of a given HBA. In the output, the Qlen field shows the queue depth loads on the HBA:

```
cli% statport -host -nodes 0 -slots 1
10:24:39 08/22/12 r/w   I/O per second       KBytes per sec  Svt ms     IOSz KB
    Port      D/C       Cur   Avg  Max    Cur    Avg    Max  Cur Avg   Cur   Avg Qlen
    0:1:1     Data   t  5113  5113 5113  67621  67621  67621 1.5 1.5  13.2  13.2    2
    0:1:2     Data   t     0     0    0      0      0      0 0.0 0.0   0.0   0.0    0
    0:1:3     Data   t     0     0    0      0      0      0 0.0 0.0   0.0   0.0    0
    0:1:4     Data   t  5844  5844 5844  80395  80395  80395 1.3 1.3  13.8  13.8    4
 ------------------------------------------------------------------------------------
       4      Data   t 10957 10957      148066 148066        1.4 1.4  13.5  13.5    6
```

If any of the HPE 3PAR StoreServ Storage ports are close to being saturated then the load on the HBA should be reduced to avoid the QUEUE FULL condition. This can be done by reducing the host LUN queue depths or presenting fewer VLUNs through these ports by moving some to other HBAs, which are less utilized.

Conclusion

Storage administrators should carefully consider the number of hosts connected to an HPE 3PAR StoreServ Storage port and the number of LUN exports. Performance degradation can occur if a target port is oversubscribed or if the host LUN queue depth values are set too low. However, if a target port is undersubscribed then performance may be improved by increasing the host LUN queue depth.

Learn more at

hp.com/go/3parstoreserv

4AA4-5094ENW, November 2015, Rev. 3