Qsan Document - White Paper

Implement iSCSI Multipath in RHEL6.5

Version 1.0
September 2014

Qsan Technology, Inc.

4F., No.103, Ruihu St.,
Neihu Dist., Taipei City 114,
Taiwan (R.O.C.)

Tel: +886-2-7720-2118
Fax: +886-2-7720-0295

Email: Sales@QsanTechnology.com
Website: www.QsanTechnology.com

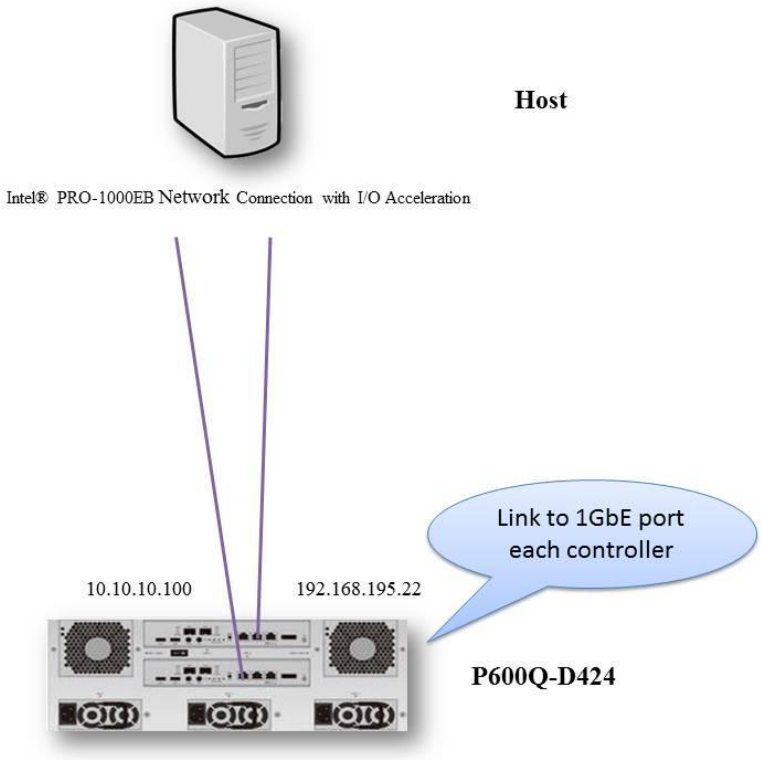
Introduction

In this document, it describes how to connect an iSCSI target in RHEL via software iSCSI initiator and use the device-mapper-multipath package to create a multipath device which is presented by Qsan P600Q-D424 via iSCSI. All Qsan iSCSI dual controller models can be used within this configuration.

Environment

Host OS:	Red Hat Enterprise Linux 6.5
Storage:	QSAN P600Q-D424
Controller firmware:	V3.2.2(201405221100)
RAM:	4GB DDR3
iSCSI data port:	10.10.10.100; 192.168.195.22
RPM Packages	iscsi-initiator-utils-6.2.0.873-10.el6.x86_64 device-mapper-1.02.79-8.el6.x86_64 device-mapper-multipath-0.4.9-72.el6.x86_64

Diagram



Installation

Before configuring the iSCSI multipath, you have to install the following rpm packages and source files (.tar.gz), so that the iSCSI service could run smoothly and without any compatible issues. Here is the order to install the packages we need:

- iscsi-initiator-utils-6.2.0.873-10.el6.x86_64.rpm
- device-mapper-1.02.79-8.el6.x86_64.rpm
- device-mapper-multipath-0.4.9-72.el6.x86_64.rpm

All the necessary rpm packages can be found in the RHEL6.5 DVD, Install them as follows:

```
# rpm -ivh /media/”RHEL_6.5 x86_64 Disc 1”/Packages/iscsi-initiator-utils-6.2.0.873-10.el6.x86_64.rpm
]# rpm -ivh /media/”RHEL_6.5 x86_64 Disc 1”/Packages/device-mapper-1.02.79-8.el6.x86_64.rpm
# rpm -ivh /media/”RHEL_6.5 x86_64 Disc 1”/Packages/device-mapper-multipath-0.4.9-72.el6.x86_64.rpm
```

Configuration

Usage of iSCSI initiator

The iSCSI initiator name can be specified in the configuration file /etc/iscsi/initiatorname.iscsi.

```
# vi /etc/iscsi/initiatorname.iscsi
InitiatorName = Your_initiator_name
```

Edit the configuration file of iSCSI initiator in /etc/iscsi/iscsid.conf, the iscsi session timeout value has to be changed to a proper value. The default value is 120 seconds, but it is too long to keep the I/O wait before the path is judged as fail and it may cause the I/O failure. Please set a shorter and proper timeout value in this configuration file.

```
# vi /etc/iscsi/iscsid.conf
node.session.timeo.replacement_timeout = 30
(Please set a proper timeout value)
```

In /etc/iscsi/iscsid.conf, it also provides others settings, such as:

```
# vi /etc/iscsi/iscsid.conf
node.startup = Automatic
(Set auto-login when discover target)
node.session.auth.authmethod = CHAP
(Enable CHAP auth)
node.session.auth.username = username
(Set CHAP username)
node.session.auth.password = password
(Set CHAP password)
```

Please restart the iSCSI service to make these changes work.

```
# service iscsi restart
```

The rpm package iscsi-initiator-utils provides a command line tool called iscsiadm. It can manage the connections to iSCSI target. The iscsiadm tool has three operational modes - discovery, node, and session. The following will introduce these modes.

1. Discovery the all port and target name by # iscsiadm -m discovery.
   Operational mode -discovery is used to discover the target, the usage is
   # iscsiadm -m discovery -t st -p target_ip

```
# iscsiadm -m discovery -t st -p 10.10.10.100
192.168.1.1:3260,0 iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.ctr1
192.168.2.1:3260,0 iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.ctr1
10.10.10.100:3260,1 iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.ctr1
192.168.4.1:3260,1 iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.ctr1
# iscsiadm -m discovery -t st -p 192.168.195.22
192.168.5.1:3260,0 iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.ctr2
192.168.6.1:3260,0 iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.ctr2
192.168.195.22:3260,1 iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.ctr2
192.168.8.1:3260,1 iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.ctr2
```

2. Users can login and logout by # iscsiadm -m node with the ip and target name.
   Operational mode -node is used to login/logout, the usage is
   # iscsiadm -m node -T target_iqn -p target_ip -l

 # iscsiadm -m node -T target_iqn -p target_ip –u

```
# iscsiadm -m node -T iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.ctr1 -p
10.10.10.100 -l
(login 10.10.10.100)
# iscsiadm -m node -T iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.ctr2 -p
192.168.195.22 -l
(login 192.68.195.22)
# iscsiadm -m node -T iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.ctr1 -p
10.10.10.100 -u
(logout 10.10.10.100)
# iscsiadm -m node -T iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.ctr2 -p
192.168.195.22 -u
(logout 192.168.195.22)
```

3. Query the list of nodes, the usage is

   # iscsiadm -m node

```
# iscsiadm -m node
192.168.1.1:3260,0 iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.ctr1
192.168.2.1:3260,0 iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.ctr1
10.10.10.100:3260,1 iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.ctr1
192.168.4.1:3260,1 iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.ctr1
192.168.5.1:3260,0 iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.ctr2
192.168.6.1:3260,0 iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.ctr2
192.168.195.22:3260,1 iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.ctr2
192.168.8.1:3260,1 iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.ctr2
```

4. If users want to clear the node list, the usage is

   # iscsiadm -m node -0 delete

5. This command will list the connected iSCSI session, it can be expressed as

   # iscsiadm -m session

```
# iscsiadm -m session
tcp: [3] 10.10.10.100:3260,1 iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.c
tr1
tcp: [4] 192.168.195.22:3260,1 iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev
```

```
0.ctr2
```

6. In session mode, the iSCSI session can be logout, the usage is

   # iscsiadm -m session -r session_id –u

```
# iscsiadm -m session -r 3 -u
Logging out of session [sid: 3, target: iqn.2004-08.com.qsantechnology:p600q-d316-000901d0
0:dev0.ctr1, portal: 10.10.10.100]
Logout of [sid: 3 target: iqn.2004-08.com.qsantechnology:p600q-d316-000901d00:dev0.ctr1, p
ortal: 10.10.10.100,3260]: successful
```

7. To log out all sessions, the usage is

   # iscsiadm -m session –u

How to setup DM-Multipath

The procedures of setup a multipath DM-Multipath are on the following.

1. To enable mpathconf, and then enable multipath support.

```
# mpathconf -h
usage: /sbin/mpathconf <command>

Commands:
Enable: --enable
Disable: --disable
Set user_friendly_names (Default n): --user_friendly_names <y|n>
Set find_multipaths (Default n): --find_multipaths <y|n>
Load the dm-multipath modules on enable (Default y): --with_module <y|n>
start/stop/reload multipathd (Default n): --with_multipathd <y|n>
chkconfig on/off multipathd (Default y): --with_chkconfig <y|n>
# mpathconf --enable
(It will create multipath.conf file as the configuration of multipath)
# service multipthd start
(To enable multipath)
```

How to exclude local disks

There are two ways that the local disks can be excluded when generating multipath devices.

1. Determine which WWN of local disks will be ignored. In this example, using the command multipath can find out the WWN of local disk /dev/sda

```
# multipath -F
(Clear all multipath device maps)
# multipath
(Create multipath)
create: mpatha(1ATA    ST31000528AS                    9V)undef ATA,ST31000528A
[size=932G feature='0' hwhandler='0' wp=undef
'-+- policy='round-robin 0' prio=1 status=undef
'- 2:0:0:0  sda8:0   undef ready running
create: mpathb (3203300137890ad00) undef Qsan,p600-d316
[size=500g feature='0' hwhandler='0' wp=undef
|-+- policy='round-robin 0' prio=1 status=undef
| '- 12:0:0:0 sdb 8:16 undef ready running
'-+- policy='round-robin 0' prio=1 status=undef
'- 13:0:0:0 sdc 8:32 undef ready running
```

TIP: The device A as follow means failover. And another one means round-robin.

A. |-+- policy='round-robin 0' prio=1 status=undef
   | '- 12:0:0:0 sdb 8:16 undef ready running
   '-+- policy='round-robin 0' prio=1 status=undef
   '- 13:0:0:0 sdc 8:32 undef ready running

B. |-+- policy='round-robin 0' prio=1 status=active
   '- 12:0:0:0 sdb 8:16 active ready running
   '- 13:0:0:0 sdc 8:32 active ready running

The WWN of local disk /dev/sda is in the parenthesis followed by the word "mpatha".

2. Edit /etc/multipath.conf, and insert the WWN of local disk into the blacklist.

```
# vi /etc/multipath.conf
blacklist {
wwid 1ATA    ST31000528AS                    9V
}
```

TIP: If you change the value of multipath.conf, you must restart multipath to take effect.

```
# service multipthd restart
```

3. User can also change the find_multipths to block the local disk

```
# multipath -find_multipaths y

OR
# vi /etc/multipath.conf
defaults{
find_multipaths yes
}
```

Next, the alias of iSCSI device will be created. The alias name will help iSCSI device to be identified easily. Find the UUID of iSCSI device in Red below:

```
# multipath -ll
mpathb (32033001378901d00) dm-3 Qsan,p600-d316
[size=500g feature='0' hwhandler='0' wp=rw
|-+- policy='round-robin 0' prio=1 status=active
| '- 12:0:0:0 sdb 8:16 active ready running
'-+- policy='round-robin 0' prio=1 status=enabled
'- 13:0:0:0 sdc 8:32 active ready running
```

1. Edit the /etc/multipath.conf again:

```
# vi /etc/multipath.conf
multipaths {
        multipath {
                wwid 32033001378901d00
                alias qsan
                path_grouping_policy multibus
#                path_checker direction
                (This line may cause multipath be invalid in different device)
                path_selector "round-robin 0"
                failback manual
```

```
        rr_weight priorities
        no_path_retry 5
}
```

2. Save the configuration file, and confirm that the persistent name to iSCSI device has been created.

```
# multipath -ll
qsan (32033001378901d00) dm-3 Qsan,p600-d316
[size=500g feature='1 queue_if_no_path' hwhandler='0' wp=ro
|-+- policy='round-robin 0' prio=1 status=active
'- 12:0:0:0 sdb 8:16 active ready running
'- 13:0:0:0 sdc 8:32 active ready running
# ls -l /dev/mapper
total 0
crw-rw---- 1 root root  10, 58 jul 28 18:34 control
lrwxrwxrwx 1 root root       7 jul 28 18:34 qsan -> ../dm-3
lrwxrwxrwx 1 root root       7 jul 28 18:34 VolGroup00-lv_home -> ../dm-2
lrwxrwxrwx 1 root root       7 jul 28 18:34 VolGroup00-lv_root -> ../dm-0
lrwxrwxrwx 1 root root      7 jul 28 18:34 VolGroup00-lv_swap -> ../dm-1
```

TIP: Usually it uses the command multipath to manage the multipath devices. Here is the parameter manual.

multipath	Without parameters, create the devmaps for the multipath devices.
-h	Print this usage text.
-l	Show multipath topology. (sysfs and DM info)
-ll	Show multipath topology. (maximum info)
-f	Flush a multipath device map.
-F	Flush all multipath device maps.
-c	Check if advice should be a path in a multipath device.
-q	Allow queue_if_no_path whenmultipathd is not running.
-d	Dry run, do not creat or update devmaps.
-r	Force devmap reload.
-p	Policy failover|multibus|group_by_serial|group_by_prio.
-b fil	Bindings file location.
-p pol	Force all maps to specified path grouping policy:

failover	1 path per priority group
multibus	all paths in 1 priority group
group_by_serial	1 priority group per serial
group_by_prio	1 priority group per priority level
group_by_node_name	1 priority group per target node

-v lvl	Verbosity level: 0 no output 1 print created devmap names only 2 default verbosity 3 print debug information
Dev	Action limited to: Multipath named 'dev' (ex: mpath0) or Multipath whose wwidis 'dev' (ex:60051..) Multipath including the path named 'dev' (ex: /dev/sda) Multipath including the path with maj:min 'dev' (ex:8:0)

Practice

Here is an example of how to create a multipath device and change the MPIO policy between failover and round-robin.

```
# multipath -F
(Clear existing device-maps)
# multipath
(Create devicemaps)
creat: qsan (32033001378901d00) undef Qsan,p600-d316
[size=500g feature='0' hwhandler='0' wp=undef
|-+- policy='round-robin 0' prio=1 status=undef
'- 12:0:0:0 sdb 8:16 undef ready running
'- 13:0:0:0 sdc 8:32 undef ready running
# fdisk -l
Disk /dev/sda: 1000.2 GB, 1000204886016 bytes
255 heads, 63 sectors/track, 121601 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk identifier: 0x000da7c4

Device Boot      Start         End      Blocks   Id  System
/dev/sda1   *           1          64      512000   83  Linux
Partition 1 does not end on cylinder boundary.
/dev/sda2              64      121602   976248832   8e  Linux LVM

Disk /dev/sdb: 107.4 GB, 107374182400 bytes
```

```
255 heads, 63 sectors/track, 13054 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk identifier: 0xa3872e75

Device Boot      Start         End      Blocks   Id  System
/dev/sdb1               1       13054   104854528    7  HPFS/NTFS

Disk /dev/sdc: 107.4 GB, 107374182400 bytes
255 heads, 63 sectors/track, 13054 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk identifier: 0xa3872e75

Device Boot      Start         End      Blocks   Id  System
/dev/sdc1               1       13054   104854528    7  HPFS/NTFS

Disk /dev/mapper/qsan: 107.4 GB, 107374182400 bytes
255 heads, 63 sectors/track, 13054 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk identifier: 0xa3872e75

# multipath -l
(List the multipath topology, default MPIO policy is multibus)
qsan (32033001378901d00) dm-3 Qsan,p600-d316
 [size=500g feature='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='round-robin 0' prio=0 status=active
 '- 12:0:0:0 sdb 8:16 active ready running
'- 13:0:0:0 sdc 8:32 active ready running
```

You can change policy (ex:failover, multibus) by editing multipath.conf

```
# vi /etc/multipath.conf
```

```
multipath {
        wwid 32033001378901d00
        alias qsan
  path_grouping_policy multibus        (You can change failover by this line)
#        path_checker directio
        path_selector "round-robin 0"
        failback manual
        rr_weight priorities
        no_path_retry 5
}
```

Now, the abstract device /dev/mapper/Qsandisk which groups /dev/sdb and /dev/sdc are ready to be used.

Mount the file system on iSCSI device at boot time(Optional)

Format the iSCSI device as EXT4 file system.

```
# mkfs.ext4 /dev/mapper/qsan
```

In order to mount a file system that exits on an iSCSI device connected through the open-iSCSI software initiator, you need to add a line to the /etc/fstab file.

```
# vi /etc/fstab
/dev/mapper/QsanDisk/mnt/DataCoreSAN ext4_netdev,defaults 0 0
```

The _netdev option will delay the time when mounting the file system on listed devices until the network has been started. Also ensures that the file system is unmounted before stopping the network at shutdown.

Performance

Test P600Q-D424 performance with 5 WD2003FYYS-02W0B0 SATA disks. And the configure setting is as follow.

- bs=1M
- Count=20480

```
# dd if=/dev/zero of=/dev/mapper/qsan bs=2M count=20480
(I/O configure setting)
20480+0 records in
20480+0 records out
42949672960 bytes (43 GB) copied, 187.942 s, 229 MB/s
```

```
# iostat -m 10
Linux 2.6.32-431.el6.x86_64 (localhost.localdomain)  08/09/2014  _x86_64_  (4 CPU)
avg-cpu:  %user   %nice %system %iowait  %steal   %idle
          0.03    0.00   17.12   36.39    0.00   46.46

Device:            tps   MB_read/s    MB_wrtn/s    MB_read    MB_wrtn
sda               0.60        0.00         0.00          0          0
dm-0              0.60        0.00         0.00          0          0
dm-1              0.00        0.00         0.00          0          0
dm-2              0.00        0.00         0.00          0          0
sdb             217.00        0.00       108.50          0       1085
sdc             218.30        0.00       109.15          0       1091
dm-3            435.30        0.00       217.65          0       2176
```

Summary

Follow the procedures; it's easy to implement iSCSI and multipath I/O in RHEL6.5. Take advantage of these two convenient packages cooperate with QSAN P series controllers, enjoy a fault-tolerance and performance enhancement technique.

Applies To

- All Qsan iSCSI dual controller models.

Reference

- RHEL 6.5 DM Multipath
  https://access.redhat.com/documentation/en-US/Red_Hat_Enterprise_Linux/6/html/6.5_Release_Notes/index.html

Obsolete

- Qsan White Paper
  QWP200801-P150C-Implement_iSCSI_multipath_in_RHEL5.pdf

Revision History

Date	Version	Owner	Description
2014/8/14	0.1	Arthur Tsai	Draft.
2014/9/18	1.0	Wilson Fang	Review.