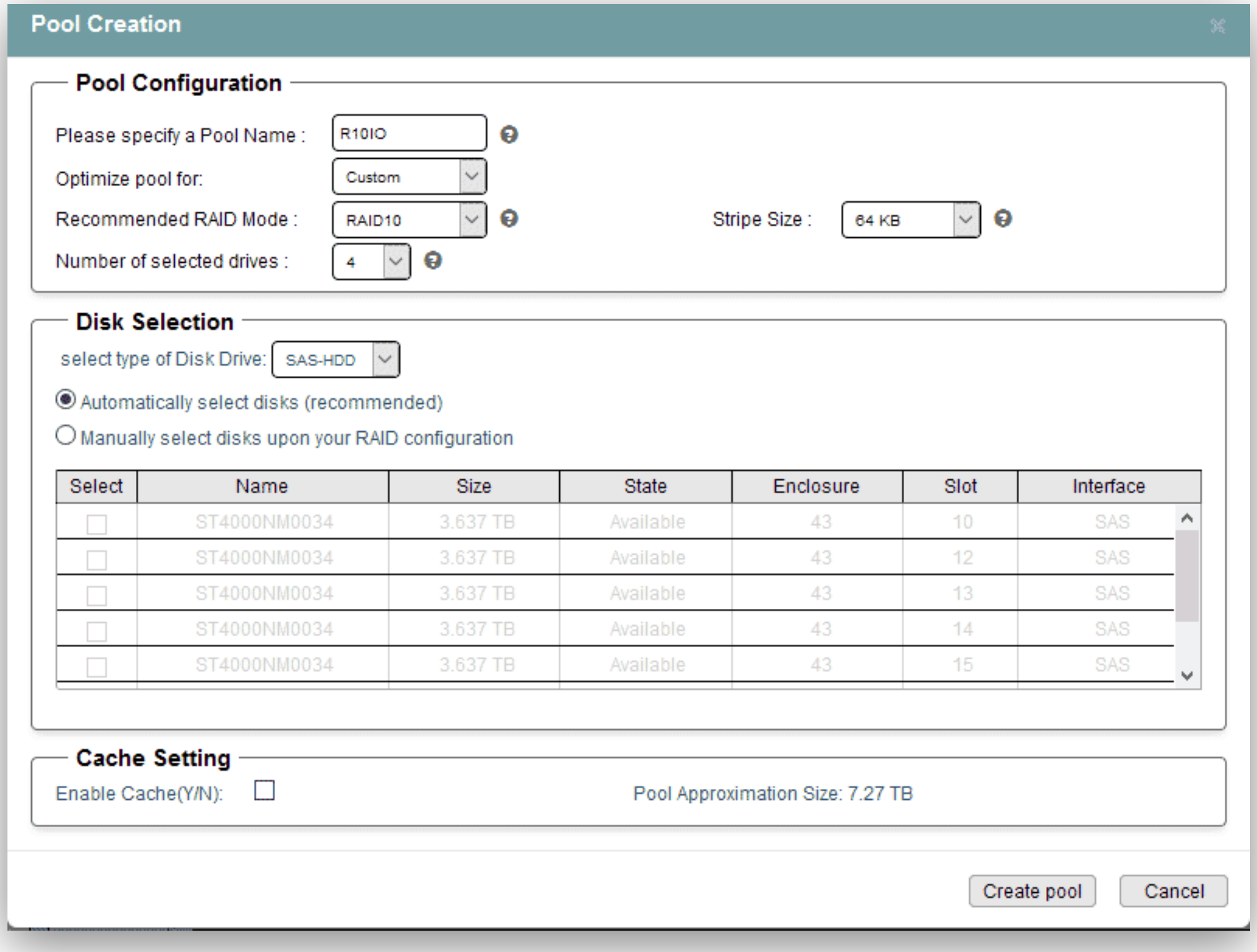
Nima Mohammadi and **Sepand Haghighi**
Prof. Asadi
9/24/1395

Advanced Storage Systems

Homework #4

1) **[Storage->Pool]** First we create the pool with the parameters given in the assignment. We do not activate the caching mechanism yet.



Pool Creation

Pool Configuration

Please specify a Pool Name : R10IO

Optimize pool for: Custom

Recommended RAID Mode : RAID10

Stripe Size : 64 KB

Number of selected drives : 4

Disk Selection

select type of Disk Drive: SAS-HDD

● Automatically select disks (recommended)

○ Manually select disks upon your RAID configuration

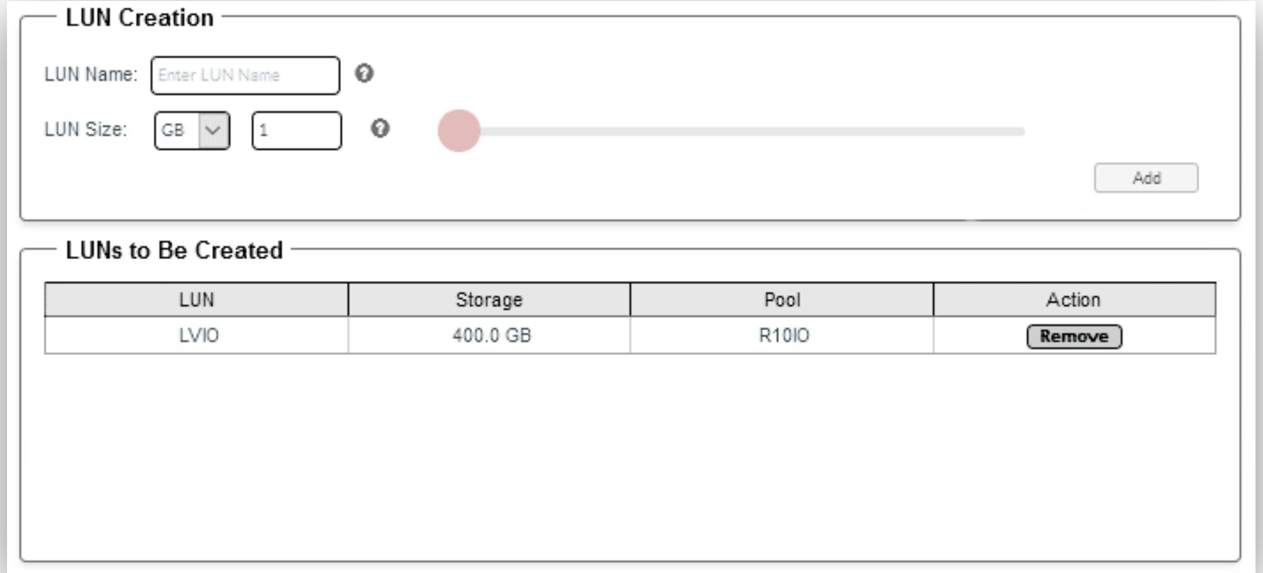
Select	Name	Size	State	Enclosure	Slot	Interface
☐	ST4000NM0034	3.637 TB	Available	43	10	SAS
☐	ST4000NM0034	3.637 TB	Available	43	12	SAS
☐	ST4000NM0034	3.637 TB	Available	43	13	SAS
☐	ST4000NM0034	3.637 TB	Available	43	14	SAS
☐	ST4000NM0034	3.637 TB	Available	43	15	SAS

Cache Setting

Enable Cache(Y/N): ☐

Pool Approximation Size: 7.27 TB

2) **[Storage->LUN]** Then we create a new LUN of 400 GB in R10IO pool.



LUN Creation

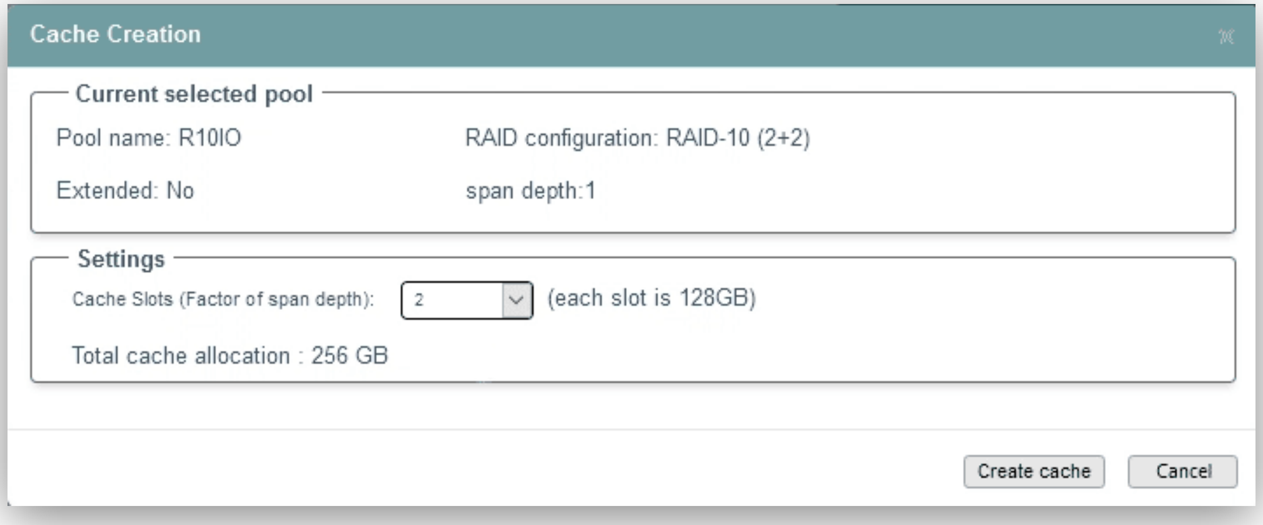
LUN Name: Enter LUN Name

LUN Size: GB 1

LUNs to Be Created

LUN	Storage	Pool	Action
LVIO	400.0 GB	R10IO	Remove

3) **[Storage->RapidStore]** Then we assign two cache slots, equal to 256 GB of SSD capacity, to pool R10IO.



Cache Creation

Current selected pool

Pool name: R10IO

RAID configuration: RAID-10 (2+2)

Extended: No

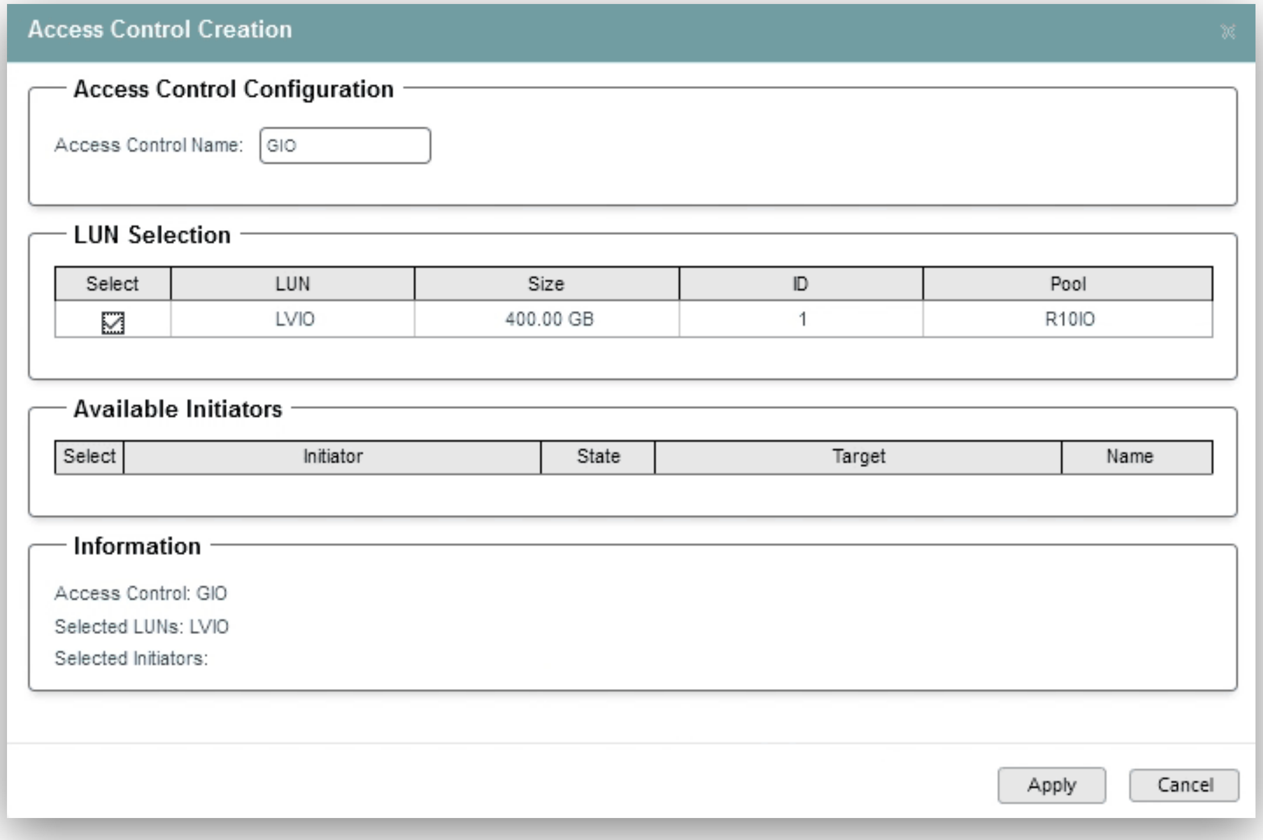
span depth:1

Settings

Cache Slots (Factor of span depth): 2 (each slot is 128GB)

Total cache allocation : 256 GB

4-A) **[Host List->Access Control]** We add an Access Control containing LUN LVIO and name it GIO.



Access Control Creation

Access Control Configuration

Access Control Name: GIO

LUN Selection

Select	LUN	Size	ID	Pool
☑	LVIO	400.00 GB	1	R10IO

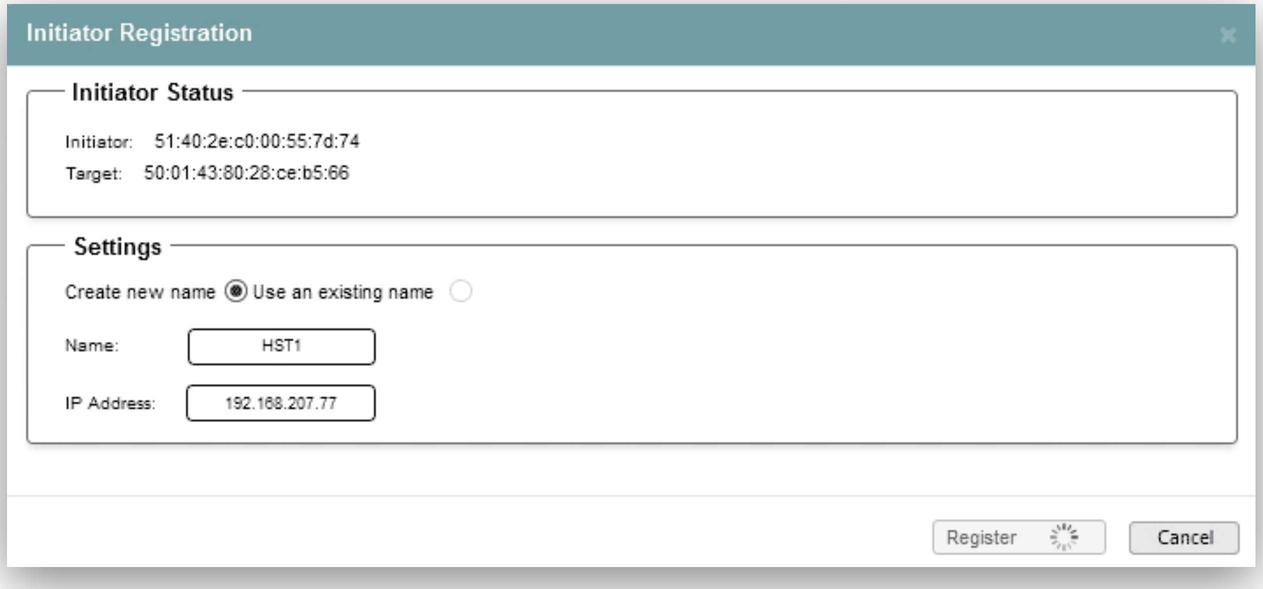
Available Initiators

Select	Initiator	State	Target	Name

Information

Access Control: GIO
Selected LUNs: LVIO
Selected Initiators:

4-B) **[Host List->Host Info]** We add a new host and set the to HST1.



Initiator Registration

Initiator Status

Initiator: 51:40:2e:c0:00:55:7d:74
Target: 50:01:43:80:28:ce:b5:66

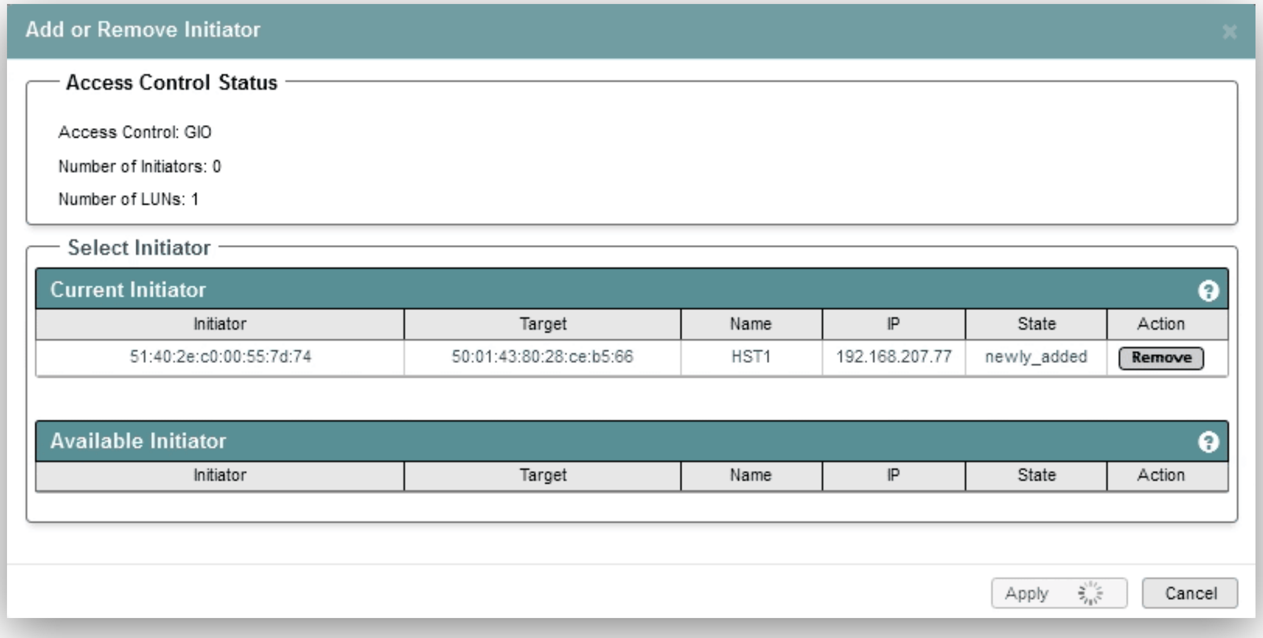
Settings

Create new name ◉ Use an existing name ○

Name: HST1

IP Address: 192.168.207.77

[Host List->Access Control] We go back and modify Access Control by adding new initiator.



Add or Remove Initiator

Access Control Status

Access Control: GIO

Number of Initiators: 0

Number of LUNs: 1

Select Initiator

Current Initiator

Initiator	Target	Name	IP	State	Action
51:40:2e:c0:00:55:7d:74	50:01:43:80:28:ce:b5:66	HST1	192.168.207.77	newly_added	Remove

Available Initiator

Initiator	Target	Name	IP	State	Action

5) Now we connect to the host using SSH. The first thing we do is to rescan scsi hosts:

```
[root@INIT1 mohammadi]# echo "- - -" > /sys/class/scsi_host/host0/scan
[root@INIT1 mohammadi]# echo "- - -" > /sys/class/scsi_host/host1/scan
[root@INIT1 mohammadi]# echo "- - -" > /sys/class/scsi_host/host2/scan
[root@INIT1 mohammadi]# echo "- - -" > /sys/class/scsi_host/host3/scan
[root@INIT1 mohammadi]# echo "- - -" > /sys/class/scsi_host/host4/scan
```

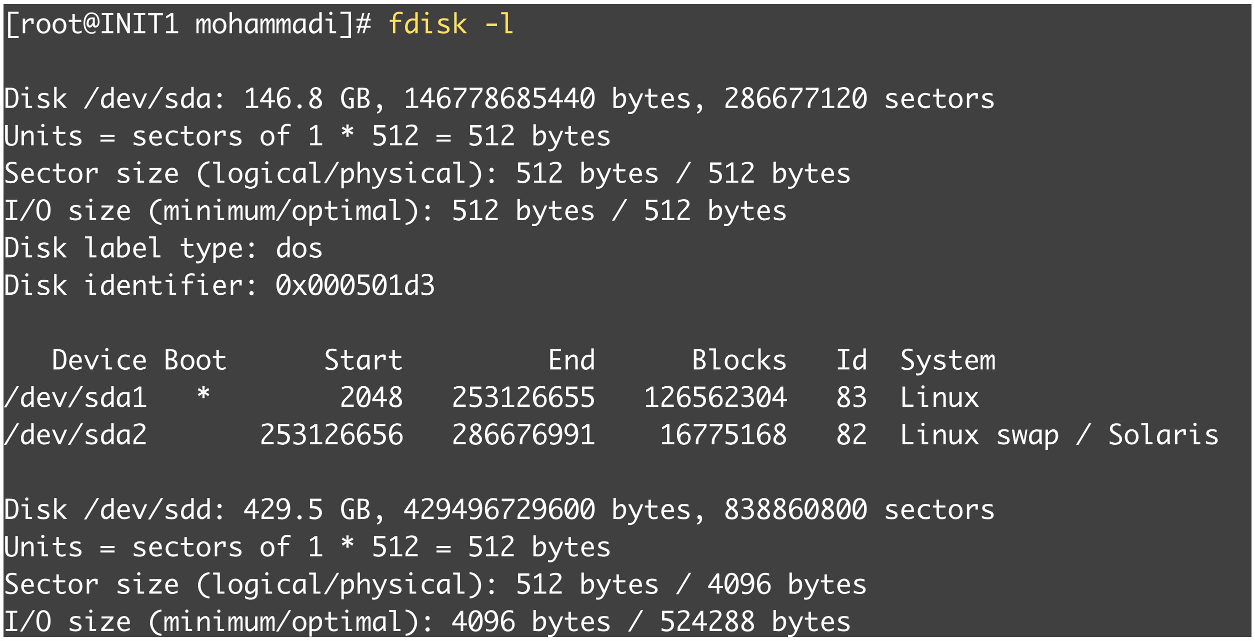
We inspect the list of SCSI devices with `lssci`:

```
[root@INIT1 mohammadi]# lsscsi
[0:0:0:0]    cd/dvd  TSSTcorp CD-ROM  TS-L162C N204  /dev/sr0
[2:0:0:0]    disk    HP       LOGICAL VOLUME   1.18  /dev/sda
[2:3:0:0]    storage HP       P400             1.18  -
[3:0:0:1]    disk    SAB      LVIO              302  /dev/sdd
[3:0:0:2]    disk    SAB      LVIO              302  /dev/sdc
```

It shows us two disks named LVIO. Running the command below removes /dev/sdc which had apparently added by the previous group:

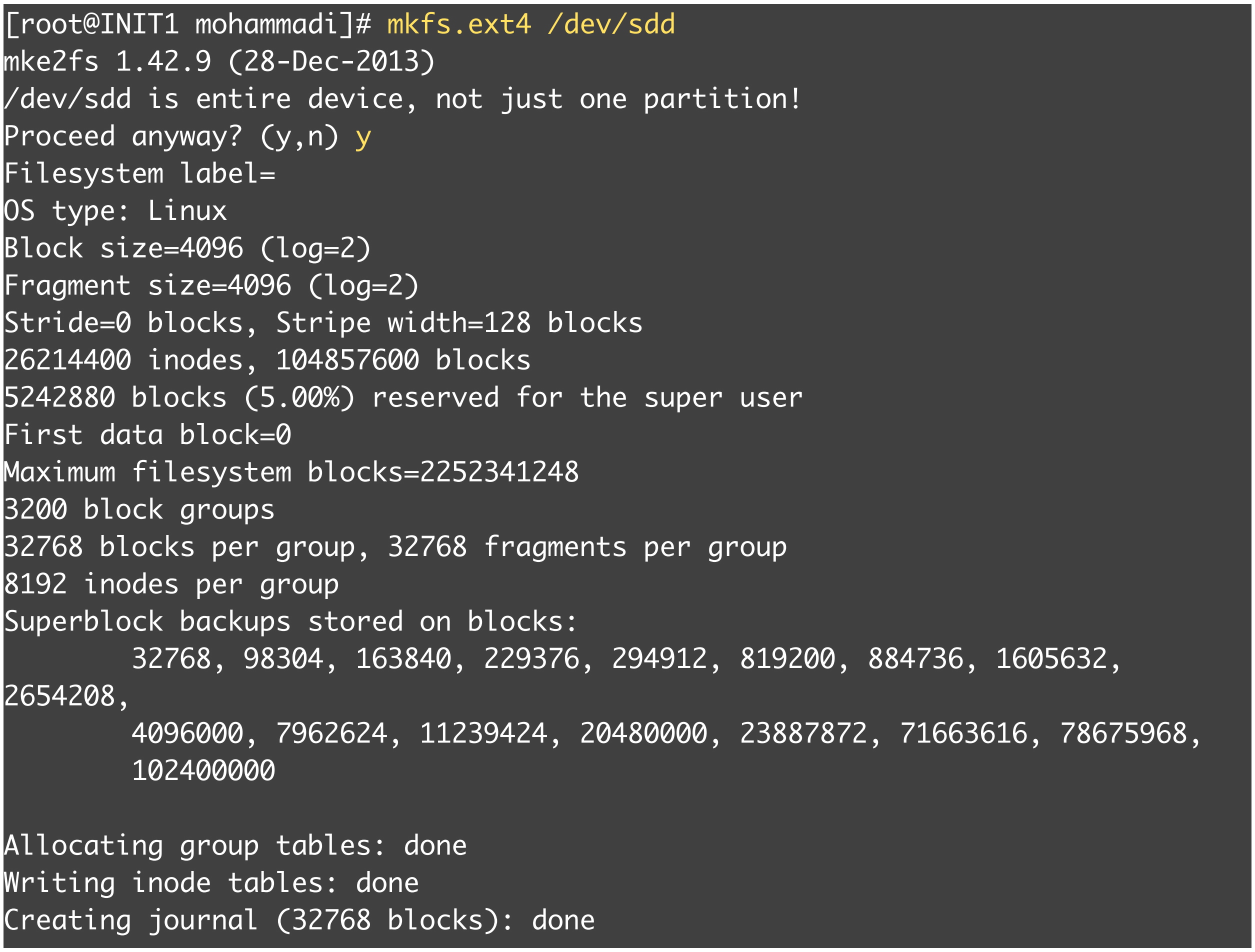
```
[root@INIT1 ~]# rescan-scsi-bus.sh -l -w -r
```

FDISK lists the disks currently attached to the system along their partitions:



```
[root@INIT1 mohammadi]# fdisk -l

Disk /dev/sda: 146.8 GB, 146778685440 bytes, 286677120 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk label type: dos
Disk identifier: 0x000501d3

   Device Boot      Start         End      Blocks   Id  System
/dev/sda1   *        2048   253126655   126562304   83  Linux
/dev/sda2       253126656   286676991    16775168   82  Linux swap / Solaris

Disk /dev/sdd: 429.5 GB, 429496729600 bytes, 838860800 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 4096 bytes
I/O size (minimum/optimal): 4096 bytes / 524288 bytes
```

6-A) Then we format the disk with Ext4 filesystem. We choose not to partition the disk:



```
[root@INIT1 mohammadi]# mkfs.ext4 /dev/sdd
mke2fs 1.42.9 (28-Dec-2013)
/dev/sdd is entire device, not just one partition!
Proceed anyway? (y,n) y
Filesystem label=
OS type: Linux
Block size=4096 (log=2)
Fragment size=4096 (log=2)
Stride=0 blocks, Stripe width=128 blocks
26214400 inodes, 104857600 blocks
5242880 blocks (5.00%) reserved for the super user
First data block=0
Maximum filesystem blocks=2252341248
3200 block groups
32768 blocks per group, 32768 fragments per group
8192 inodes per group
Superblock backups stored on blocks:
        32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632,
2654208,
        4096000, 7962624, 11239424, 20480000, 23887872, 71663616, 78675968,
        102400000

Allocating group tables: done
Writing inode tables: done
Creating journal (32768 blocks): done
```

```
Writing superblocks and filesystem accounting information: done
```

6-B) We proceed with mounting the newly-formatted disk:

```
[root@INIT1 mohammadi]# mount /dev/sdd /home/mohammadi/mntpoint/
```

6-C) To test the mounted LUN, we use dd to write from /dev/zero and after a few second kill the process:

```
[root@INIT1 mntpoint]# dd if=/dev/zero of=test.dd
^C998762+0 records in
998761+0 records out
511365632 bytes (511 MB) copied, 3.86728 s, 132 MB/s
```

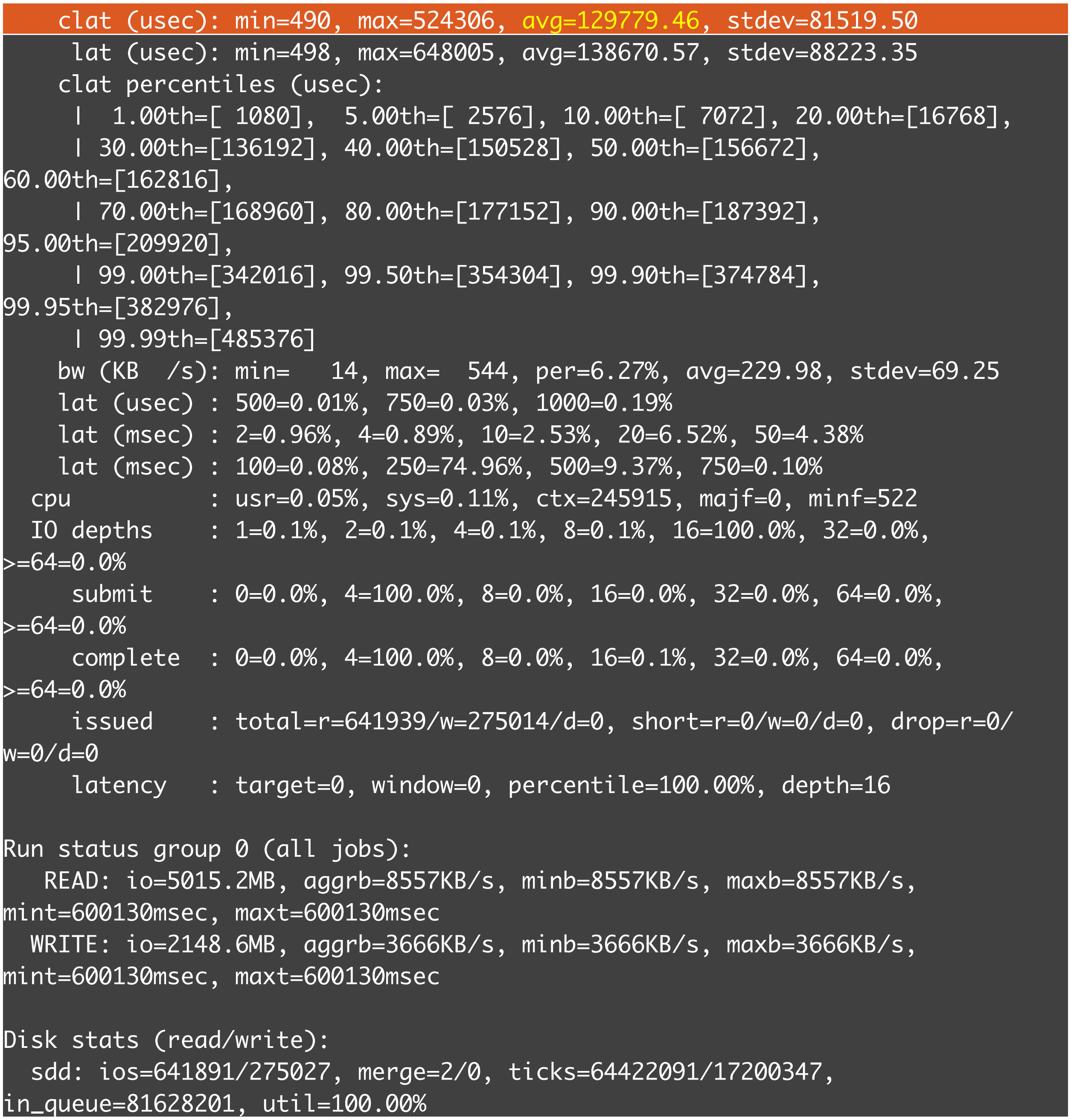
It reports the writing speed of 132 MB/s.

7) The command below uses fio to evaluate the performance of the storage:

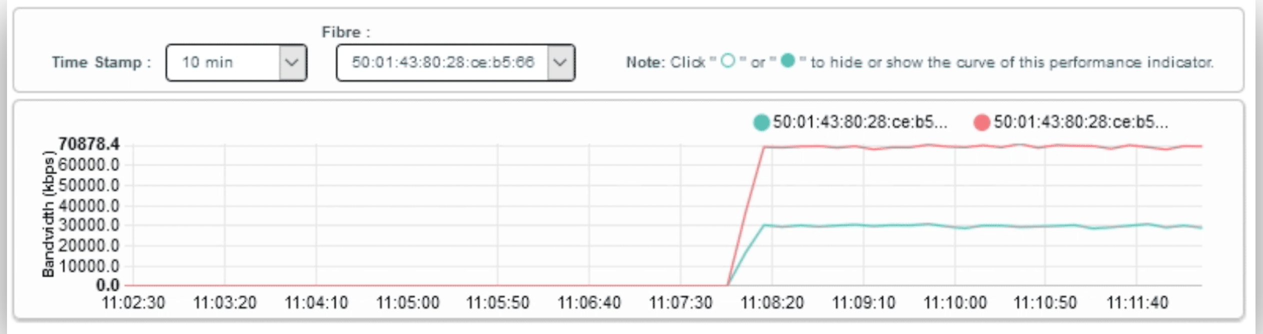
```
[root@INIT1 FIO]# fio --filename=/dev/sdd --direct=1 --rw=randrw --
rwmixread=70 --bs=8k --ioengine=libaio --iodepth=16 --numjobs=16 --
runtime=600 --refill_buffers --randrepeat=0 --random_distribution=random --
norandommap --time_based --group_reporting --name=sdd_Test –
output=sdd.fio.out
```

Following is output of `fio` for this specific task:

```
[root@INIT1 FIO]# cat sdd.fio.out
sdd_Test: (g=0): rw=randrw, bs=8K-8K/8K-8K/8K-8K, ioengine=libaio, iodepth=16
...
fio-2.2.8
Starting 16 processes

sdd_Test: (groupid=0, jobs=16): err= 0: pid=21061: Wed Dec 14 22:38:15 2016
  read : io=5015.2MB, bw=8557.4KB/s, iops=1069, runt=600130msec
    slat (usec): min=5, max=245925, avg=8770.53, stdev=34140.19
    clat (usec): min=561, max=730835, avg=171117.49, stdev=77037.23
     lat (usec): min=580, max=789166, avg=179888.60, stdev=84192.13
    clat percentiles (msec):
     |  1.00th=[   11],  5.00th=[   19], 10.00th=[   38], 20.00th=[  149],
     | 30.00th=[  155], 40.00th=[  161], 50.00th=[  167], 60.00th=[  172],
     | 70.00th=[  180], 80.00th=[  188], 90.00th=[  310], 95.00th=[  334],
     | 99.00th=[  367], 99.50th=[  383], 99.90th=[  510], 99.95th=[  529],
     | 99.99th=[  578]
    bw (KB  /s): min=  198, max=  912, per=6.29%, avg=537.98, stdev=109.33
  write: io=2148.6MB, bw=3666.6KB/s, iops=458, runt=600130msec
    slat (usec): min=5, max=241793, avg=8890.53, stdev=34367.75
```



```
    clat (usec): min=490, max=524306, avg=129779.46, stdev=81519.50
     lat (usec): min=498, max=648005, avg=138670.57, stdev=88223.35
    clat percentiles (usec):
     |  1.00th=[ 1080],  5.00th=[ 2576], 10.00th=[ 7072], 20.00th=[16768],
     | 30.00th=[136192], 40.00th=[150528], 50.00th=[156672],
60.00th=[162816],
     | 70.00th=[168960], 80.00th=[177152], 90.00th=[187392],
95.00th=[209920],
     | 99.00th=[342016], 99.50th=[354304], 99.90th=[374784],
99.95th=[382976],
     | 99.99th=[485376]
    bw (KB  /s): min=   14, max=  544, per=6.27%, avg=229.98, stdev=69.25
    lat (usec) : 500=0.01%, 750=0.03%, 1000=0.19%
    lat (msec) : 2=0.96%, 4=0.89%, 10=2.53%, 20=6.52%, 50=4.38%
    lat (msec) : 100=0.08%, 250=74.96%, 500=9.37%, 750=0.10%
  cpu          : usr=0.05%, sys=0.11%, ctx=245915, majf=0, minf=522
  IO depths    : 1=0.1%, 2=0.1%, 4=0.1%, 8=0.1%, 16=100.0%, 32=0.0%,
>=64=0.0%
     submit    : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.0%,
>=64=0.0%
     complete  : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.1%, 32=0.0%, 64=0.0%,
>=64=0.0%
     issued    : total=r=641939/w=275014/d=0, short=r=0/w=0/d=0, drop=r=0/
w=0/d=0
     latency   : target=0, window=0, percentile=100.00%, depth=16

Run status group 0 (all jobs):
   READ: io=5015.2MB, aggrb=8557KB/s, minb=8557KB/s, maxb=8557KB/s,
mint=600130msec, maxt=600130msec
  WRITE: io=2148.6MB, aggrb=3666KB/s, minb=3666KB/s, maxb=3666KB/s,
mint=600130msec, maxt=600130msec

Disk stats (read/write):
  sdd: ios=641891/275027, merge=2/0, ticks=64422091/17200347,
in_queue=81628201, util=100.00%
```

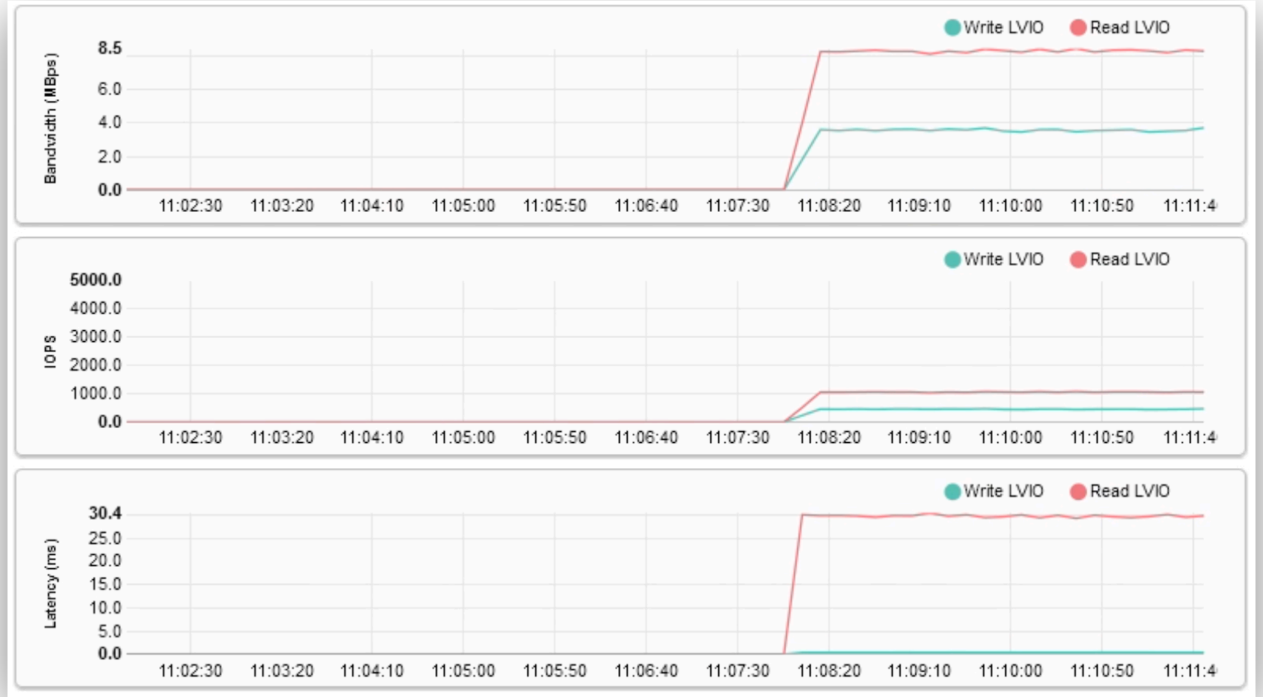


Time Stamp : 10 min

Fibre : 50:01:43:80:28:ce:b5:66

Note: Click " ○ " or " ● " to hide or show the curve of this performance indicator.

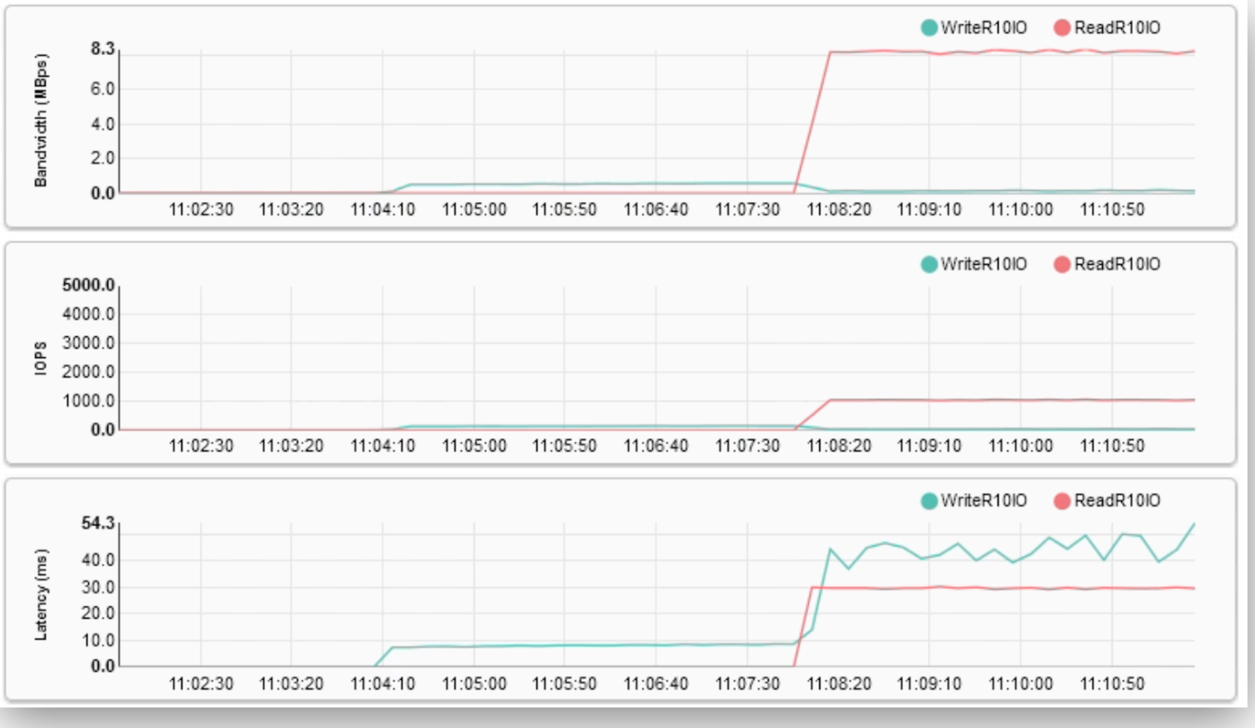
Inspecting the monitor for the fiber channel interface confirms the throughput to be of 70 to 30 ratio for read and write (8.7 MBps to 3.7 MBps), as instructed by the `fio` command `rwmixread` parameter.



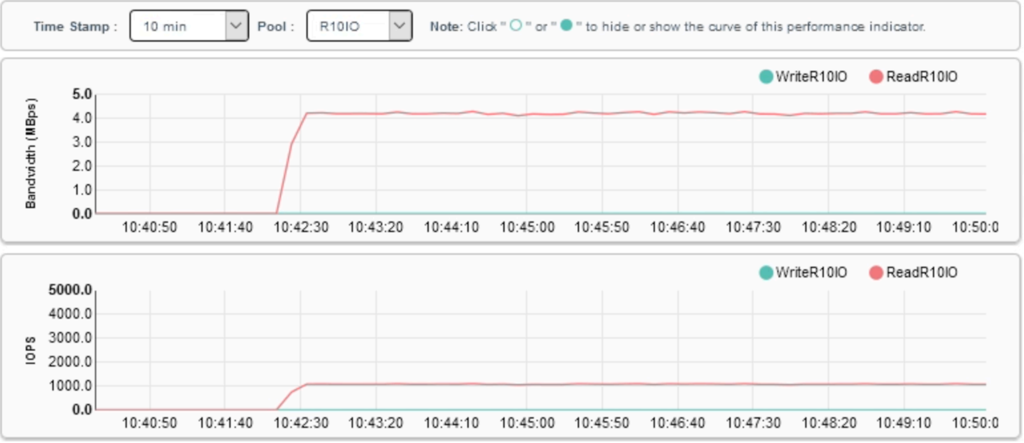
The bandwidth shown for the LUN LVIO is consistent with the bandwidth reported for fiber channel, negligibly higher for fiber channel due to packetization overhead. Moreover, bandwidth is also equal to the value reported by `fio`.

IOPS for the LUN depicted in monitoring section of panel, shown in figure above, correspond to the values reported in `fio` output, i.e. 1069 and 458 for read and write operations, respectly.

As for latency, while `fio` reports 171 ms and 129 ms for read and write operations, respectly, the monitoring section in the panel reports them to be 30ms and 0.1ms. The difference between reported values stems from the fact that fio also observes the latency caused by the fiber channel interface, while HPDS only monitors the underlying disks activity. Results suggest that the significant portion of reported latency by `fio` is caused by fiber channel. HPDS reports write latency to be zero possibly because the caching mechanism used for RapidStore is write-back. This is only our speculation as the manual for HPDS refrains to specify the actual caching mechanism used in RapidStore.

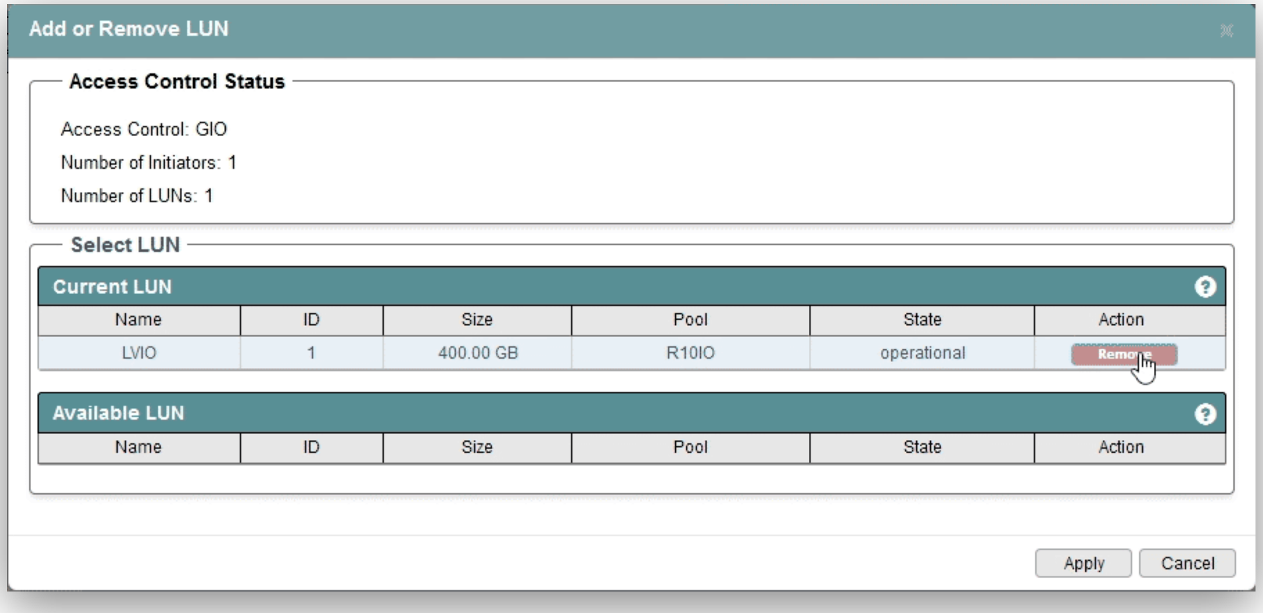


Sadly, the curves for the pool is not as realistic. Figure above motivates our claim as wrong IOPS and BW readings reported for write operations. Other tests we have performed with `fio` also resulted in inconsistent readings for write. For example, another test we conducted with 70:30 read/write ratio that as can be seen below, write operations were not captured by the monitor:



Time Stamp : 10 min Pool : R10IO Note: Click " ○ " or " ● " to hide or show the curve of this performance indicator.

8-A) **[Host List->Access Control]** We remove the the LUN from Access Control.



Add or Remove LUN

Access Control Status

Access Control: GIO

Number of Initiators: 1

Number of LUNs: 1

Select LUN

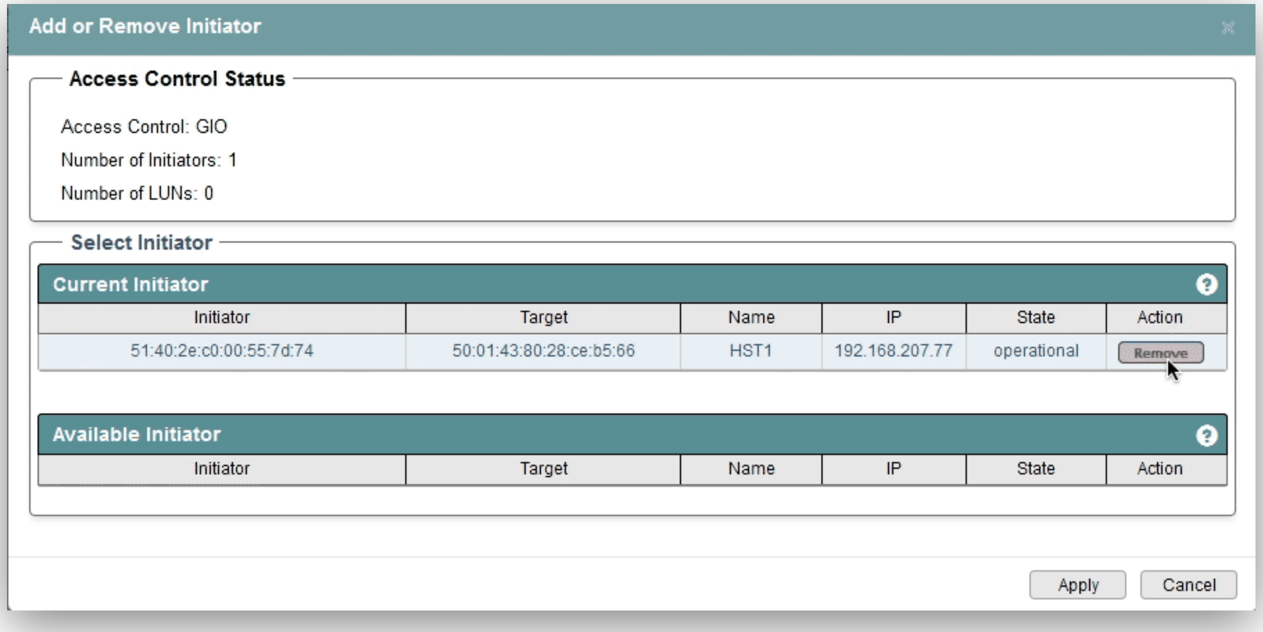
Current LUN

Name	ID	Size	Pool	State	Action
LVIO	1	400.00 GB	R10IO	operational	Remove

Available LUN

Name	ID	Size	Pool	State	Action

8-B) **[Host List->Access Control]** Now we remove the the Initiator from Access Control.



Add or Remove Initiator

Access Control Status

Access Control: GIO

Number of Initiators: 1

Number of LUNs: 0

Select Initiator

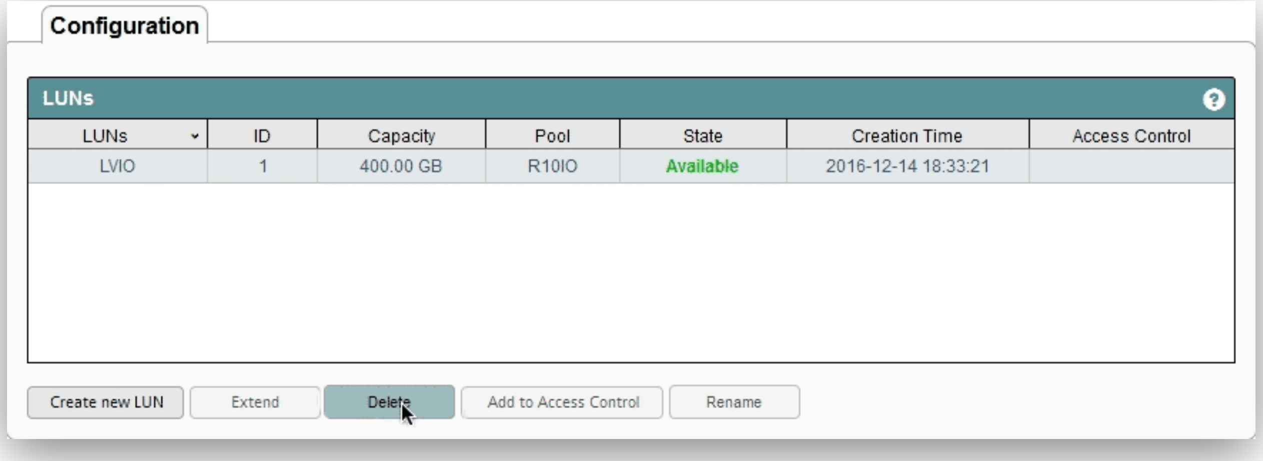
Current Initiator

Initiator	Target	Name	IP	State	Action
51:40:2e:c0:00:55:7d:74	50:01:43:80:28:ce:b5:66	HST1	192.168.207.77	operational	Remove

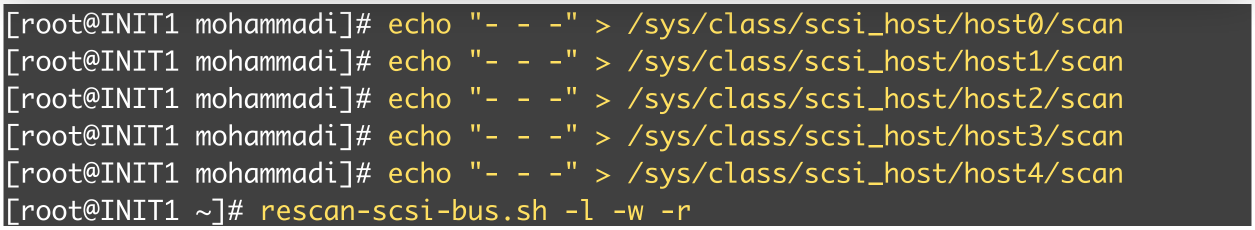
Available Initiator

Initiator	Target	Name	IP	State	Action

9) **[Storage->LUN]** We remove the LUN.
Running the commands below will rescan SCSI bus and remove the LUN. This can be confirm using `lscsi` command.

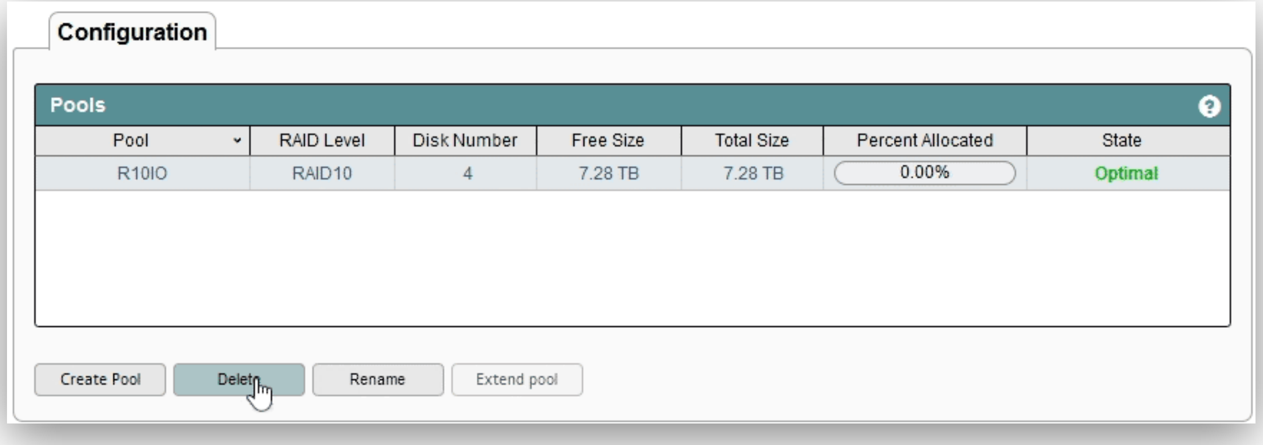


LUNs	ID	Capacity	Pool	State	Creation Time	Access Control
LVIO	1	400.00 GB	R10IO	Available	2016-12-14 18:33:21	



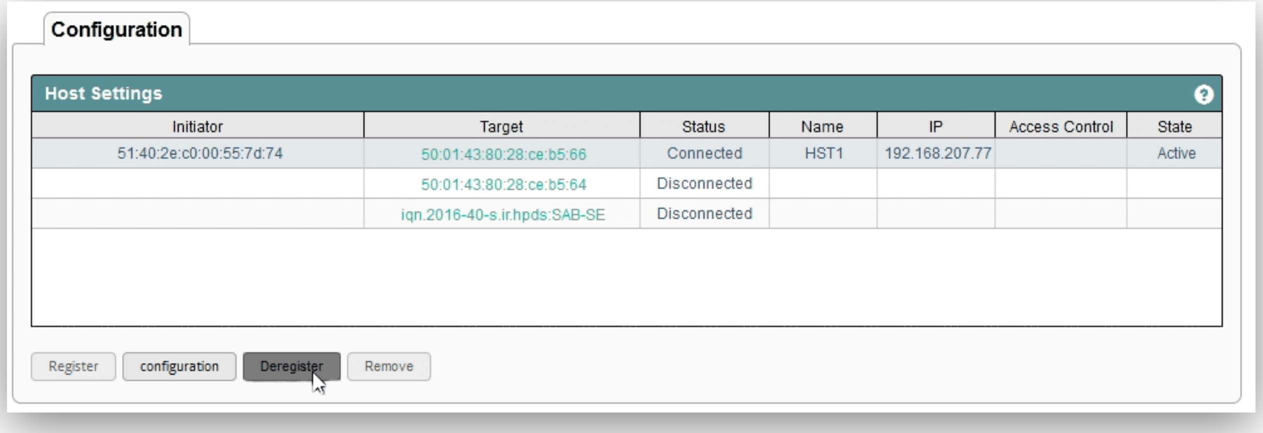
```
[root@INIT1 mohammadi]# echo "- - -" > /sys/class/scsi_host/host0/scan
[root@INIT1 mohammadi]# echo "- - -" > /sys/class/scsi_host/host1/scan
[root@INIT1 mohammadi]# echo "- - -" > /sys/class/scsi_host/host2/scan
[root@INIT1 mohammadi]# echo "- - -" > /sys/class/scsi_host/host3/scan
[root@INIT1 mohammadi]# echo "- - -" > /sys/class/scsi_host/host4/scan
[root@INIT1 ~]# rescan-scsi-bus.sh -l -w -r
```

10) **[Storage->Pool]** We proceed by removing the pool R10IO we had created.



Pool	RAID Level	Disk Number	Free Size	Total Size	Percent Allocated	State
R10IO	RAID10	4	7.28 TB	7.28 TB	0.00%	Optimal

11) **[Host List->Host Info]** And eventually, we deregister the initiator.



Initiator	Target	Status	Name	IP	Access Control	State
51:40:2e:c0:00:55:7d:74	50:01:43:80:28:ce:b5:66	Connected	HST1	192.168.207.77		Active
	50:01:43:80:28:ce:b5:64	Disconnected				
	iqn.2016-40-s.ir.hpds:SAB-SE	Disconnected				

Notes:

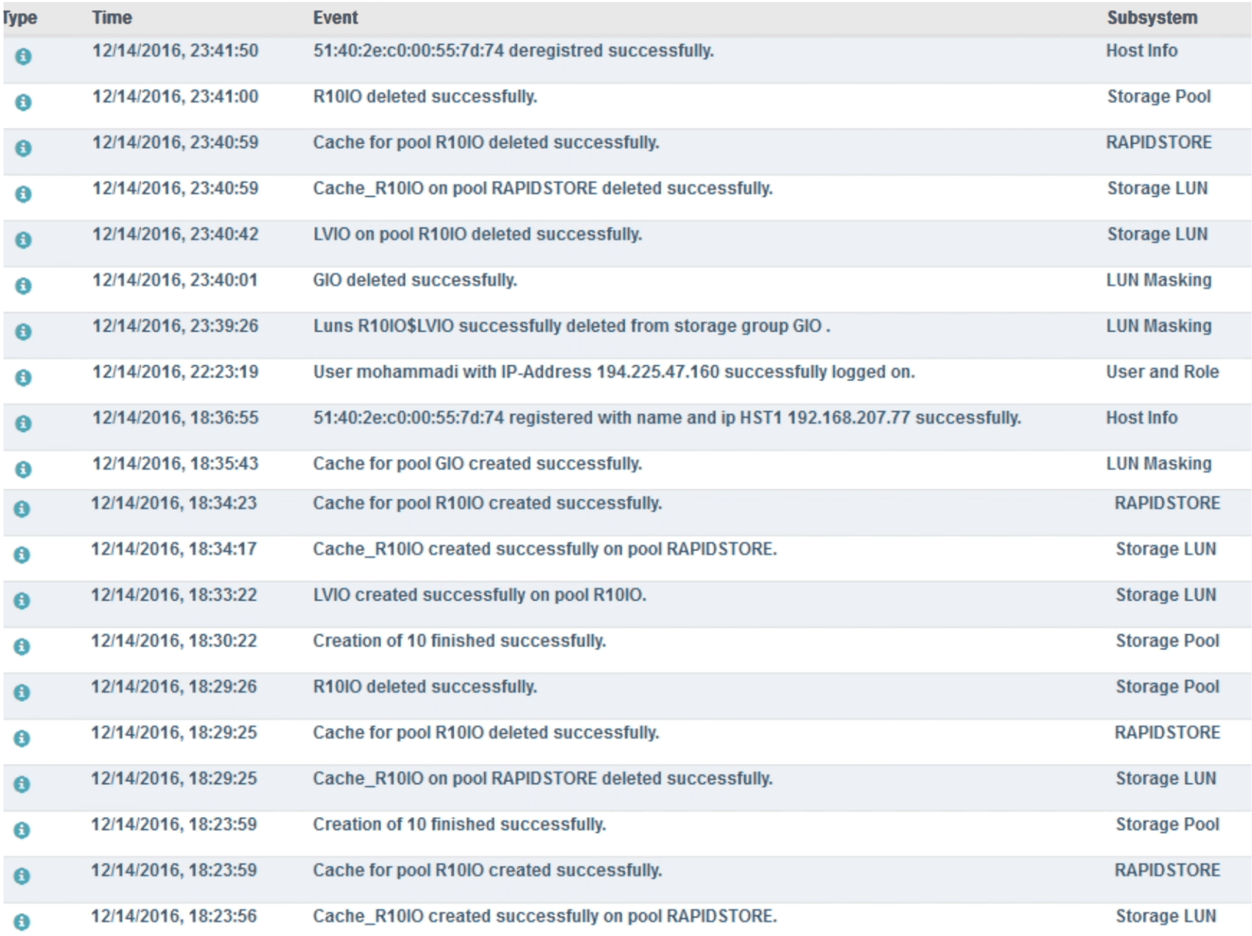
In summary, the storage system was quite stable and its mode d'emploi was surprisingly very easy.

There performance, based on the reported IOPS is acceptable for a disk-based storage system, but is not so, considering that we employed an SSD-base caching mechanism to speed things up. We have not compared the results with a configuration without caching, to better asses RapidStore, as it was not instructed by the assesment, but we expected the impact of using RapidStore to be more apparent.

As for observed issues, we stumbled upon cases where the reading from the monitor for the pool were incorrect. This is explained in more details in section 7 of this document. It would be also more informative if the user who originated and caused an event would be loggen in the events.

The dissapointing part for us was that some features claimed in product website, www.hpds.ir, are nowhere to be seen! These include instant snapshots and data encryption. It also claims no SPoF, while the storage, at its current state, does not support more than one physical node. The active backplanes and motherboards are also considered as SPoFs and for SAB to be truely without a single point of failure, the storage must be spanned over several nodes.

System Logs



Type	Time	Event	Subsystem
ℹ	12/14/2016, 23:41:50	51:40:2e:c0:00:55:7d:74 deregistred successfully.	Host Info
ℹ	12/14/2016, 23:41:00	R10IO deleted successfully.	Storage Pool
ℹ	12/14/2016, 23:40:59	Cache for pool R10IO deleted successfully.	RAPIDSTORE
ℹ	12/14/2016, 23:40:59	Cache_R10IO on pool RAPIDSTORE deleted successfully.	Storage LUN
ℹ	12/14/2016, 23:40:42	LVIO on pool R10IO deleted successfully.	Storage LUN
ℹ	12/14/2016, 23:40:01	GIO deleted successfully.	LUN Masking
ℹ	12/14/2016, 23:39:26	Luns R10IO$LVIO successfully deleted from storage group GIO .	LUN Masking
ℹ	12/14/2016, 22:23:19	User mohammadi with IP-Address 194.225.47.160 successfully logged on.	User and Role
ℹ	12/14/2016, 18:36:55	51:40:2e:c0:00:55:7d:74 registered with name and ip HST1 192.168.207.77 successfully.	Host Info
ℹ	12/14/2016, 18:35:43	Cache for pool GIO created successfully.	LUN Masking
ℹ	12/14/2016, 18:34:23	Cache for pool R10IO created successfully.	RAPIDSTORE
ℹ	12/14/2016, 18:34:17	Cache_R10IO created successfully on pool RAPIDSTORE.	Storage LUN
ℹ	12/14/2016, 18:33:22	LVIO created successfully on pool R10IO.	Storage LUN
ℹ	12/14/2016, 18:30:22	Creation of 10 finished successfully.	Storage Pool
ℹ	12/14/2016, 18:29:26	R10IO deleted successfully.	Storage Pool
ℹ	12/14/2016, 18:29:25	Cache for pool R10IO deleted successfully.	RAPIDSTORE
ℹ	12/14/2016, 18:29:25	Cache_R10IO on pool RAPIDSTORE deleted successfully.	Storage LUN
ℹ	12/14/2016, 18:23:59	Creation of 10 finished successfully.	Storage Pool
ℹ	12/14/2016, 18:23:59	Cache for pool R10IO created successfully.	RAPIDSTORE
ℹ	12/14/2016, 18:23:56	Cache_R10IO created successfully on pool RAPIDSTORE.	Storage LUN